

Sarah Djebali - PhD
INSERM CRCN Researcher (HDR)
IRSD INSERM U1220 - Bat B - CHU Purpan
31024 Toulouse Cedex 3
Email: sarah.djebali@inserm.fr
Nationality: French
Born: January 15th 1978 (46 years-old)
Married, 2 children

**Bioinformatics researcher expert in functional
genomic data analysis and method development**

I Training

- 2002-2006 **PhD in bioinformatics** at the IBENS (Institute of Biology of Ecole Normale Supérieure) and the IBISC (Computer science, Integrative Biologie and Complex systems) laboratories
A formal framework for expert eukaryotic gene annotation
With honors (*mention très honorable*)
- 2001-2002 **Research Master** of Application of Mathematics and Computer science to Biology (AMIB) from Evry University (now Paris Saclay)
Automatic generation of Tagcc programs by training
With honors (*mention bien*)
- 1998-2001 **Master in Computer science (software engineer)** from ENSEIRB (Ecole Nationale Supérieure of Electronics, Computer science and Radiocommunications of Bordeaux)
With honors (*mention bien*)

II Professional career

- From 2019 **INSERM permanent researcher (CRCN)**, Toulouse, France
IRSD (Institute of Research in Digestive Health), INSERM U1220
Complex gene expression regulation to better understand common diseases
- 2018-2019 **INRAE Senior Postdoc**, Toulouse, France
GenPhySE (Genetics, Physiology and Livestock Systems), INRAE U1289
Robust identification of structural variants in livestock species
- 2015-2018 **Agreenskills+ INRAE Senior Postdoc**, Toulouse, France
GenPhySE (Genetics, Physiology and Livestock Systems), INRAE U1289
Functional and comparative annotations of the genomes of four vertebrate livestock species
- 2010-2015 **CRG Staff Scientist**, Barcelona, Spain
CRG (Center for Genomic Regulation)
Coordination of the transcriptomic data analysis of the international ENCODE and MouseENCODE projects

2006-2010	CRG/IMIM junior Postdoc , Barcelona, Spain CRG/IMIM (Institute of Medical Research) Supervisor: Roderic Guigó <i>Characterisation of the human transcriptional landscape (ENCODE)</i>
2002-2006	PhD / Assistant Professor at Evry University , Evry, France IBISC and IBENS laboratories PhD supervisors: Franck Delaplace and Hugues Roest Crollius <i>A formal framework for expert eukaryotic gene annotation</i>
2002 (6 months)	AMIB Master internship , Evry, France IBISC laboratory Supervisor: Franck Delaplace <i>Automatic generation of Tagcc programs by training</i>
2001 (10 months)	ENSEIRB project graduation , Sheffield, U.K. Sheffield Hallam University Supervisor: Barrie W. Jervis <i>Neural network development for electroencephalogram data classification</i>
2000 (3 months)	ENSEIRB second year project , Bordeaux, France LABRI (Laboratory for research in computer science) Supervisor: David Sherman <i>Inductive databases for the study of gene neighborhood</i>

III Teaching

The following table provides a summary of my teaching activities in the reverse chronological order:

Years	Number of hours	Subject	Audience	Location
2024	2h	Functional annotation databases and their interrogation	INSERM workshop attendees	Bordeaux
2024	2h	Causal variant fine-mapping using functional annotations	INSERM workshop attendees	Bordeaux
2022	12h	Linux and the command line	Bioinformatics Master	UPS, Toulouse
2021-2022	4h	Functionnal annotation and common diseases	Bioengineering Master	UPS, Toulouse
2012-2020	72h	Theoretical and practical course about ENCODE and/or the analysis of RNA-seq data	Researchers and engineers	Pasteur/EMBO/CRG/SIB
2011	8h	Theoretical and practical computer science courses (UNIX and programming)	Researchers in Biology	CRG
2002-2006	280h	Theoretical and practical computer science and bioinformatics courses	License and Master students	Evry University

IV Research

IV.1 Publications

I have co-authored **37 original articles** in peer-reviewed journals, of which:

- 7 as first author or first co-author

- 4 as last author

I have also contributed to:

- 1 review article
- 1 book chapter (as first author)
- 1 popular article (as first author)

IV.2 Oral presentations

I have given 15 oral presentations in conferences, of which:

- 5 for which I was invited
- 10 in which my work has been selected by peer-review

IV.3 H-index and citations

According to Research Gate (September 19th 2024), my h-index is **27** and my number of citations is **29,003**, with *Landscape of transcription in human cells* cited **5,649** times. Three of my articles have also been commented by other researchers.

IV.4 Prizes

I have obtained the *Vanguardia de la Ciència* prize for my contribution to *Landscape of transcription in human cells*, *Nature*, 2012.

V Supervision

The following table indicates the people I have scientifically supervised or co-supervised by reverse chronological order:

Person's name	Project title	Level	Supervision ratio	Supervision dates
Victor Lefebvre	Divergence of duplicated genes' cis-regulatory landscape	M2	33%	January to July 2024
Océane Carpentier	Knowledge graphs to identify miRNA/gene relationships	M2	50%	January to July 2024
Martin Racoupeau	Comparison of methods for microRNA identification	M2	50%	January to July 2023
Zoé Gerber	Identification of causal variants in coronary artery disease	M2	100%	September 2022 to June 2023
Tess Azevedo	Gene co-expression networks and module identification in livestock species	M1	33%	June to October 2022
Smahane Chalabi	Impact of the maternal diet's fiber content on the transcriptomes and epigenomes of fetus and piglets	Postdoc	50%	October 2021 to March 2023
Tristan Hoellinger	Genome-wide identification of enhancer/gene relationships	Master	100%	September 2020 to February 2022
Pierre Guenzi-Tibéri	Development of a <code>Nextflow</code> pipeline for quantifying the expression of small RNAs from high-throughput sequencing (HTS) data	M1	50%	June to August 2021
Claire Martin	Combining enhancer/gene relationships and genetic data to identify new variants and genes associated to Parkinson's disease	Engineer	50%	September 2020 to March 2021
Camille Mestre	Genome-wide identification of enhancer/gene relationships	M2	60%	January to July 2018
Matthew Smart	Comparison of chromatin accessibility and chromatin structure data	L3	70%	March to September 2017
Nikki Gentle	Transcriptional Signatures of Viral Control in HIV-1 Infected South African Women	PhD student	30%	January to July 2014
Alessandra Breschi	Human and mouse transcriptome comparison	PhD student	50%	November 2012 to October 2015
Bernardo Rodríguez Martín	Chimeric transcript identification in vertebrate genomes	M2 and PhD student	100% and then 30%	January 2013 to October 2016
Emilio Palumbo	Robust pipeline development for the analysis of HTS functional data	Engineer	80%	November 2012 to October 2015
Matthieu Muffato	Development of a user interface for the <code>Exogean</code> gene identification program	M1	80%	April to July 2006
Charles Vejnar	Development of a pseudogene detection module for the <code>Exogean</code> gene identification program	M1	80%	April to July 2005
Khaled Yahiaoui	Improvement and evaluation of <code>Exogean</code> , a eukaryotic gene identification program	M2	80%	January to July 2005

VI Cooperation programs

The following table lists the cooperation programs I have been involved in and my role in them:

Funding source (Country)	Program's name	Program's goal	My role	Program's dates
H2020 (Europe)	GENE-SWitCH	The regulatory GENomE of SWine and CHicken: functional annotation during development	Co-WP leader	2019-2023
INRAE Animal Genetic's division (France)	Raffine	RNA-seq and ATAC-seq for fine-mapping QTL related to pig meat quality	Coordinator	2017-2018
Agreenskills+ (Europe)	FR-AgENCODE data analysis	Multi-species annotation of transcriptome and chromatin structure in domesticated animals	Co-coordinator	2015-2018
NHGRI (USA)	ENCODE3 transcriptome	Investigating the transcriptome of primary cells	Co-coordinator of the data analysis	2011-2015
NHGRI (USA)	MouseENCODE	Producing an Encyclopedia of mouse functional elements	Co-coordinator of the transcriptomic data analysis	2008-2012
NHGRI (USA)	ENCODE2 Data Analysis Center	Coordinating and assisting the integrative analysis of HTS functional data produced by the ENCODE consortium	Co-coordinator	2008-2012
NHGRI (USA)	ENCODE2 Transcriptome	The first exhaustive characterization of the human transcriptome landscape	Co-coordinator of the data analysis	2008-2012