Fernando **Pozo**

Bioinformatics | Machine Learning | Data Science

Based in Alcalá de Henares (Madrid)

☐ +34 697 300 806 | ✉ fpozoc@gmx.com | 🏠 fpozoc.com | 💻 fpozoc | fpozoc | 🔗 fernanpozo | 🐦 @fpozoca

Summary

I am a last year PhD candidate in Bioinformatics at Spanish National Cancer Research Center - CNIO Bioinformatics Unit. My current research interests include the application of statistical and computing techniques to solve bioinformatics and data driven problems. Before having the opportunity to be included in this fantastic group, I was improving my scientific and analytics training with some internships within different environments, including pharmaceutical industry or biotechnology research.

In my previous years as a student, I joined the study of a BSc. in Health Biology at University of Alcalá with a posterior MSc. of Bioinformatics for Health Sciences at Pompeu Fabra University and several technical specializations in statistics, programming and mathematics. This background has given me the opportunity to join the PhD in Bioinformatics at Autonomous University of Madrid. Moreover in 2020, I was certified as Software Instructor by The Carpentries, and I got the Deep Learning Professional Certificate programme by IBM.

My main current research is focused on the study of the functional impact of alternative splicing at protein level. I am also currently involved as one of the members of the GENCODE project. For the last 2 years, I have been participating as a lecturer teaching foundations of bioinformatics.

I am passionate about data driven solutions linked with new technologies, innovation and health sciences. Experienced in bioinformatics, statistics, machine learning, data visualization, software development, HPC services and Unix environments. Skilled in several programming languages for data analysis like Python (main), R or Bash, I am also familiar with frontend/backend web development, version control, cloud computing, containerization and workflow management systems.

Experience

Bioinformatician at Bioinformatics Unit
Madrid, Spain

SPANISH NATIONAL CANCER RESEARCH CENTRE (CNIO)   *Feb. 2018 - PRESENT*

- Refining the human genome annotations and participating in projects of GENCODE, a consortium which aims to annotate all evidence-based gene features (genes, transcripts, coding sequences, etc) in the entire human and mouse genomes at a high accuracy.
- Applying machine learning predictive methods to high amount of data from omics experiments.
- Bioinformatics Unit staff member. *bioinformatics.cnio.es/staff/fpozoc/*

Visiting Professor
Madrid, Spain

UNIVERSITY CARLOS III OF MADRID   *Sep. 2021 - PRESENT*

- Teaching the whole subject titled "Machine Learning in Healthcare" within the last year of B.Sc. in Data Science and Engineering

Lecturer
Madrid, Spain

FRANCISCO DE VITORIA UNIVERSITY   *Sep. 2019 - PRESENT*

- Lecturer at B.Sc. in Biotechnology (Year 4). Subject: "Advanced Bioinformatics". *gitlab.com/fpozoc/advanced_bioinformatics*

Information Technology Consultant
Remote

START BARCELONA   *Feb. 2020 - Sep. 2020*

- Web development, tech consulting, and AI Research.

Data Analyst
Madrid, Spain

KANTAR HEALTH SA   *Jul. 2017 - Feb. 2018*

- Giving support in market research projects development focus on creation, analysis and optimization of digital strategies set up by pharmaceutical industry at European level.
- Managing the backend/frontend of a pharma web platform service (PHP, SQL, HTML5/CSS/JS and Bash scripting for the maintenance of the server).
- Developing a new mobile application for a pharma aiming to control and make clinical related surveys using the PhoneGap Technology (PHP, SQL, HTML5/CSS/JS).

Machine Learning and Bioinformatics Internship at Bioinformatics Unit
Madrid, Spain

SPANISH NATIONAL CANCER RESEARCH CENTRE (CNIO)   *Sep. 2017 - Feb. 2018*

- Main project: "Machine Learning algorithm predicts the functional importance of potential Alternative Splicing".

Bioinformatics Research Internship
Madrid, Spain

NATIONAL CENTER FOR BIOTECHNOLOGY (CNB-CSIC)   *Feb. 2016 - Jul. 2016*

- Main project: Bioinformatics analysis of data from whole genome sequencing. (New workflow for whole genome analysis and some tools which improve genome annotation)
- Research Areas: Genomics (Sequence alignment, gene finding, genome assembly and genome annotation). Epidemiology (Mathematical and computational modelling of infectious diseases).

Education

Doctor of Philosophy - PhD, Bioinformatics, Year 4
Madrid, Spain

AUTONOMOUS UNIVERSITY OF MADRID
Sep. 2018 - Jun. 2021

- Study of the functionality of Alternative Spliced Isoforms.

Software Instructor
Remote

THE CARPENTRIES
Nov. 2020 - Feb. 2021

- Certificate: *fpozoc.com/static/media/docs/carpentry_certificate_fpozoc.pdf*

Deep Learning Professional Certificate
Remote

IBM
Jul. 2019 - Jul. 2020

- Courses: Applied Deep Learning Capstone Project, Deep Learning Fundamentals with Keras, Deep Learning with Python and PyTorch, Deep Learning with Tensorflow and Using GPUs to Scale and Speed-up Deep Learning. Certificate: *credentials.edx.org/credentials/ff96c76397c74061acbe17450f1c7563*

Master of Science in Bioinformatics for Health Science
Barcelona, Spain

UNIVERSITY POMPEU FABRA
Sep. 2016 - Jul. 2018

- Composed by theorethical lessons in Bioinformatics, which is an interdisciplinary field which combines mathematical and computer sciences with biology and/or medicine, and a complete Research Master Thesis in an External Centre. Average grade: 81/100.

Bachelor's Degree of Science in Health Biology
Alcala de Henares, Spain

UNIVERSITY OF ALCALA
Sep. 2013 - Jun. 2016

- Strong foundation in Biology, Chemistry, Statistics, quantitative methods, Genetics, Cellular and Structural Biology, Epidemiology and Human Health. Average grade: 73/100.

Volunteer

Member of RSG Spain - Madrid Node
Madrid, Spain

ISCB STUDENT COUNCIL
Oct. 2018 - PRESENT

- I am actively participating in meetings and activities promoted by RSG Spain Madrid Node. (http://www.rsg-spain.iscbsc.org/)

Member of CNIOSA (CNIO Student Association)
Madrid, Spain

SPANISH NATIONAL CANCER RESEARCH CENTRE (CNIO)
Sep. 2018 - PRESENT

- Responsible for carrying out and maintaining the website design and contents and Slack Administrator.

Member
International

ISCB - INTERNATIONAL SOCIETY FOR COMPUTATIONAL BIOLOGY
Aug. 2018 - PRESENT

- Being a student member of this organization I participated at 17TH European Conference On Computational Biology (ECCB 2018 - Athens)

Speaker at 8th LabDay
Madrid, Spain

SPANISH NATIONAL CANCER RESEARCH CENTRE (CNIO)
Dec. 2018

- Speaker Selected for a Short Talk representing the Bioinformatics Unit in the CNIO Traditional Day.

Publications

Peripheral Blood Mononuclear Cells Predict Therapeutic Efficacy of Immunotherapy in NSCLC
Cancers

ROGADO ET AL.
May. 2022

DOI: 10.3390/cancers14122898

APPRIS: selecting functionally important isoforms
NAR

RODRIGUEZ ET AL.
Jan. 2022

DOI: 10.1093/nar/gkab1058

APPRIS principal isoforms and MANE Select transcripts in clinical variant interpretation
bioRxiv

POZO ET AL.
Sep. 2021

PMID: 34046593 PMCID: PMC8140736 DOI: 10.1093/nargab/lqab044

The clinical importance of tandem exon duplication-derived substitutions
Nucleic Acids Research

MARTINEZ-GOMEZ ET AL.
Jul. 2021

PMID: 34302486 PMCID: PMC8373072 DOI: 10.1093/nar/gkab623

Assessing the functional relevance of splice isoforms
NAR Genomics and Bioinformatics

POZO ET AL.
Jun. 2021

DOI: 10.1101/2021.09.17.460749

GENCODE 2021 *Nucleic Acid Research*

Frankish et al. *Jan. 2021*

PMID: 33270111 PMCID: PMC7778937 DOI: 10.1093/nar/gkaa1087

An analysis of tissue-specific alternative splicing at the protein level *PLOS Computational Biology*

Rodriguez et al. *Oct. 2020*

PMID: 33017396 PMCID: PMC7561204 DOI: 10.1371/journal.pcbi.1008287

Few SINEs of life: Alu elements have little evidence for biological relevance despite elevated translation *NAR Genomics and Bioinformatics*

Martinez-Gomez et al. *Dic. 2019*

PMID: 31886458 PMCID: PMC6924539 DOI: 10.1093/nargab/lqz023

GENCODE reference annotation for the human and mouse genomes *Nucleic Acid Research*

Frankish et al. *Jan. 2019*

PMID: 30357393 PMCID: PMC6323946 DOI: 10.1093/nar/gky955

Posters & Symposiums

TRIFID: determining functional isoforms *Remote*

ISMB 2020 *Jul. 2021*

- Evolution and Comparative Genomics COSI: poster and video presentation. *gitlab.com/fpozoc/ismb2020_evocompgen_826*

Pathogenic Mutations and Functional Importance *Madrid, Spain*

Progress Report Talk at Spanish National Cancer Research Centre (CNIO) *Oct. 2020*

- This speech was based on the methodology and results obtained by the moment about the pathogenic mutations annotated in databases as ClinVar and its relationship with Alternative Splicing protein isoforms.

Machine Learning Predicts the Functional Importance of Alternative Splicing Isoforms *Madrid - Spain*

ONCONET-SUDOE Workshop: The patient journey: Information Technologies focused on the cancer patient *Apr. 2019*

- Poster session presented at the ONCONET-SUDOE Workshop on Innovative IT for healthcare.

Madrid - Spain

CNIO 8TH Lab Day *Dec. 2018*

- Group selected talk included in the program of the annual CNIO Presentations Day.

Madrid - Spain

CNIC PhDay 2018 *Nov. 2018*

- Poster session presented at the 17th Centro Nacional de Investigaciones Cardiovasculares Carlos III (CNIC).

Granada - Spain

JBI 2018 & VI RSG Spain Bioinformatics Student Symposium *Nov. 2018*

- Talk included in the program of the Student Symposium. The speech focused to explain my project to another students. Poster session presented at the XIV SYMPOSIUM ON BIOINFORMATICS (JBI 2018)

More Than 90% of Alternative Proteins are Not Important *Madrid, Spain*

Progress Report Talk at Spanish National Cancer Research Centre (CNIO) *Oct. 2018*

- This speech was based on the methodology and results obtained by the moment about the impact of alternative splicing on the Human Proteome. It was found that the alternative proteins we do detect have certain characteristics and we have developed a Machine Learning approach which allows us to predict the importance of Alternatively Spliced Isoforms.

Fewer Than 10% of Annotated Alternative Splice Isoforms Appear to be Functionally Important *Athens, Greece*

ECCB 2018 *Sept. 2018*

- Poster session presented at the 17th European Conference on Computational Biology.

Projects

A system for the identification and interpretation of pulmonary pathologies from Chest X-rays images using Deep Learning *Madrid - Spain*

UC3M *Jun. 2022*

- I coordinated the Bachelor's thesis of one of my students from the Machine Learning in Healthcare UC3M subject. This project intended to create an end-to-end web application system that combines Deep Learning techniques like NLP, transfer learning, or image analysis to facilitate the interpretation and identification of pulmonary diseases from Chest-X-rays images.

ONCOTHON - Big Data & Cancer Genomics Challenge

Granada - Spain

JBI 2018

Nov. 2018

- I took part of the team awarded with the 1st prize in the category of Big Data & Cancer Genomics proposing an integrated workflow to improve Personalized Treatment exploiting Cancer vulnerabilities.

Roche D3 Challenge

Barcelona, Spain

Freelance

Apr. 2017 - Jun. 2017

- An automatic way to perform a complete docking and protein-ligand system building was created as a project of Molecular Simulations. It compares the results of 3 docking algorithms selecting the best ligand pose for each of them. The program was written in Python and developed for the data set of small molecules provided by Roche in the D3 Challenge contest.

VAHO - Official Website Development

Madrid, Spain

Freelance

Mar. 2017 - PRESENT

- Development of the Spanish music band VAHO website. Simple and responsive webpage developed with HTML5/CSS/JS and a PHP form. At the moment I am working in SEO and analytics techniques of digital marketing in order to improve the social media impact. *vahomusica.com/*

CMDC.py - Predictive tool of residues interactions from correlated mutations and distances correlations

Barcelona, Spain

University Pompeu Fabra

Jan. 2017 - Apr. 2017

- Complete tool which calculates correlated mutations and distance correlations to predict aminoacid interactions of a given protein. It also contains standalone scripts of mutual information calculation and contact map. Programmed in Python and Bash.

GAMENDER - Application based on games finder

Barcelona, Spain

University Pompeu Fabra

Jan. 2017 - Mar. 2017

- Web application developed in HTML5/CSS/JS as Front-End Programming languages and MySQL and PHP database management, as back-end languages development. It provides you any game released in the last 20 years if you have doubts about what game you want to play.

Modeling Brain Regions with a Predictive Naïve Bayes Model

Barcelona, Spain

University Pompeu Fabra

Jun. 2017 - PRESENT

- Machine learning approach. Predictive Naïve Bayes Model classifier which try to predict brain regions from its subsequent Splicing Events in Human Genome (GTEX consortium). Programmed in Python and Bash.

Bionformatics Analysis of data from Whole Genome Sequencing

Madrid, Spain

University of Alcala

Jun. 2017 - PRESENT

- It describes the use of several selected tools according to criteria of speed and efficiency and an automatic pipeline of Whole Genome Sequencing in Bacteria. The analysis of *Mycobacterium*, *E. Coli* and *Citrobacter* were detailed in order to establish an automatic resource which could provide knowledge for the researcher.

Languages

Programming

Main, Python

Over 3000 lines, Bash, R

Over 1500 lines, HTML5, CSS, Perl

Familiar, SQL, PHP, JavaScript

Preferred libraries, pandas, numpy, PyTorch, scipy, scikit-learn, keras, fastai, xgboost, lightgbm, biopython, altair, seaborn, (Python), Bioconductor, ggplot2, shiny (R), JQuery (JS)

Spoken and Written

Spanish, Native Language

English, Advanced Level (C1)

French, Reading Fluency. Experienced with clients.

Skills

Personally, I am experienced in both academic and business work environment, communicating project records, technical concepts and analytics-driven findings, I usually feel great interest for challenges, also having an innovative spirit.

Technically, I have strong data and analytical skills, knowledge about machine learning (both supervised and unsupervised wide range of techniques), statistical modeling, inferential statistics or database management.

In general, I have had the opportunity to lead and execute bioinformatics and data science projects, ability to work in remote environments, handling large and/or complex data sets, and keeping a close contact with technologies like containerization (Docker, Singularity), workflow management (Snakemake, Nextflow), cloud computing (AWS, Google Cloud), server management (Slurm), prototyping tools (Jupyter, RStudio), Linux and OS X, visualization tools, relational database architectures, Unix Shell tools, full stack web development, Git, ViM, R Markdown, LaTeX or GSuite tools.

Courses and Extracurricular Activity

2019	**Statistical Learning**, Standford Lagunita	*Remote*
2018	**Machine Learning Crash Course**, Google	*Remote*
2018	**Python Programming Bible	Networking, GUI, Email, XML, CGI**, Udemy	*Remote*
2017	**The Complete Web Developer Course 2.0**, Udemy	*Remote*
2016	**Python Programming**, National Center for Biotechnology (CNB-CSIC)	*Madrid, Spain*
2016	**Methods in Biostatistics**, National Center for Biotechnology (CNB-CSIC)	*Madrid, Spain*
2016	**Digital Marketing**, IAB Spain	*Madrid, Spain*
2015	**Intensive Course in English language immersion of Life and Health**, Menendez Pelayo University	*Santander, Spain*
2013	**Biomathematics**, Complutense University of Madrid	*Madrid, Spain*
2013	**International Excellence Fellowship. CEI BioTic**, University of Granada	*Granada, Spain*

References

Michael Tress, PhD, Staff Scientist at CNIO - *mtress@cnio.es*

Fatima Al-Sharour, PhD, Head of Bioinfromatics Unit at CNIO - *falshahrour@cnio.es*

Marcos Cámara-Donoso, Machine Learning Engineer at Lifebit - *marcos@lifebit.ai*

Isaac Filella Mercè, PhD, Postdoctoral Researcher at BSC - *isaac.filella1@bsc.es*

Gabriel Piedrafita Fernández, PhD, Professor at UCM - *gpiedraf@ucm.es*