José Guilherme de Almeida

MSc, computational biologist (he/him)

✉ josecpa@ebi.ac.uk   🆔 0000-0002-1887-0157

○ github.com/josegcpa   🌐 josegcpa.ml

Education

Oct 2017 – Ongoing	**European Bioinformatics Institute** and **University of Cambridge** *PhD in Computational Biology*
Sep 2015 – Sep 2017	**Universidade de Coimbra** *MSc in Cell and Molecular Biology* with honours
Sep 2012 – Jul 2015	**Universidade de Coimbra** *BSc in Biochemistry*

Relevant Research Experience

2017 - Ongoing	**Cancer data science group @ EMBL-EBI** Suvervised by Moritz Gerstung and George S. Vassiliou *PhD fellow* • Development of machine- and deep-learning methods to detect and characterize vast collections of cells in digitalised whole blood slides in a haematological cancer context. Predictive modelling of disease genotype using machine-learning methods to uncover cytomorphological profiles • Statistical and Bayesian modelling of longitudinal targeted sequencing experiments to uncover the genetic and non-genetic factors driving clonal expansion. Phylogenetic and phylodynamic modelling of the lifelong trajectories of clones using single-cell colonies in healthy individuals
2016 - 2017	**Data-driven molecular design group @ CNC-UC** Supervised by Irina S. Moreira *MSc student* • Development of machine-learning protocols to determine hot-spots (important residues) in the binding interfaces of proteins • Structural and statistical analysis of large collections of protein-protein complexes and structural characterization of complexes with no known structure

Technical skills

Programming	**Python** (advanced user), **R** (advanced user), **C** (beginner)
Machine-learning	**Machine-learning** - `scikit-learn` (Python), `caret` (R) **Deep-learning** - `tensorflow`, `pytorch` (Python)
Computer-vision	`scikit-image`, `OpenCV` (Python)
Statistical analysis	**Frequentist methods** **Bayesian methods** (particularly MCMC)
Data visualization	`ggplot2` (R)
Workflows	**Containerisation** (`Docker`) **Workflow management** (`snakemake`)

Professional certificates

2021 - ongoing	**Colorado Boulder University** *Statistical Inference for Data Science Applications Specialization (online)*
2021	**Erasmus University Rotterdam** *Econometrics course (online)*

Teaching Experience

2019	**EMBL Heidelberg** *2019 EMBL Lautenschlager Summer School* Teaching young graduate students about practical bioimage analysis
2016	**Universidade de Coimbra** *Workshops on Introductory Programming* Teaching young students about programming in Python and R

Fellowships and honors

2017	**Universidade de Coimbra** *Merit fellowship for exceptional curricular performance during my MSc*
2017 - ongoing	**European Molecular Biology Laboratory** *PhD fellowship*

Peer-reviewed research grants

2018-2021	**Fundação para Ciências e Tecnologia** Deep learning in cancer drug discovery: a pipeline for the generation of new therapies *Role: team member*
2018-2021	**Fundação para Ciências e Tecnologia** Membrane proteins development of new computational approaches and its application to GPCRs *Role: team member*

Other activities

2016 - 2017	**Junior Enterprise for Science and Technology (JEST)** *Co-founder* JEST is a junior initiative I founded with a few colleagues that is dedicated to data science training among young students and services to external businesses
2018	**20th EMBL PhD Symposium** *Organization, speaker contact* Contacted different high-profile researchers to invite them to present at the 20th EMBL PhD Symposium
2019	**EBI-Sanger-Cambridge PhD Symposium (eSCAMPS) 2019** *Website design* Developed and designed the website for the 2019 eSCAMPS

Languages

Portuguese	**Native**
English	**Proficient**
Spanish	**Beginner**

Scientific works

2021
1. Fabre, M. A., **de Almeida, J. G.**, Fiorillo, E., Mitchell, E., *et al.* The longitudinal dynamics and natural history of clonal haematopoiesis. *bioRxiv (under review at Nature)* (2021).

2. Preto, A. J., Matos-Filipe, P., **de Almeida, J. G.**, Mourão, J. & Moreira, I. S. in *Artificial Neural Networks* 267–288 (Springer, 2021).

2020
3. Preto, A. J., Barreto, C. A., Baptista, S. J., **de Almeida, J. G.**, *et al.* Understanding the Binding Specificity of G-Protein Coupled Receptors toward G-Proteins and Arrestins: Application to the Dopamine Receptor Family. *Journal of Chemical Information and Modeling* **60,** 3969–3984 (2020).

2019
4. R Magalhães, P., Machuqueiro, M., **Almeida, J. G.**, Melo, A., *et al.* Dynamical rearrangement of human epidermal growth factor receptor 2 upon antibody binding: effects on the dimerization. *Biomolecules* **9,** 706 (2019).

2018
5. Lemos, A., Melo, R., Preto, A. J., **Almeida, J. G.**, *et al.* In silico studies targeting G-protein coupled receptors for drug research against Parkinson's disease. *Current neuropharmacology* **16,** 786–848 (2018).

6. Melo, R., Lemos, A., Preto, A. J., **Almeida, J. G.**, *et al.* Computational approaches in antibody-drug conjugate optimization for targeted cancer therapy. *Current topics in medicinal chemistry* **18,** 1091–1109 (2018).

7. Melo, R., Lemos, A., Preto, A. J., Bueschbell, B., *et al.* An Overview of Antiretroviral Agents for Treating HIV Infection in Paediatric Population. *Current medicinal chemistry* (2018).

8. Preto, A. J., **Almeida, J. G.**, Schaarschmidt, J., Xue, L. C., *et al.* Computational Tools for the Structural Characterization of Proteins and Their Complexes from Sequence-Evolutionary Data. *Encyclopedia of Analytical Chemistry: Applications, Theory and Instrumentation,* 1–19 (2018).

2017
9. **Almeida, J. G.**, Preto, A. J., Koukos, P. I., Bonvin, A. M. & Moreira, I. S. Membrane proteins structures: A review on computational modeling tools. *Biochimica et Biophysica Acta (BBA)-Biomembranes* **1859,** 2021–2039 (2017).

10. Bastos, F. C., Corceiro, V. N., Lopes, S. A., **de Almeida, J. G.**, *et al.* Effect of tolbutamide on tetraethylammonium-induced postsynaptic zinc signals at hippocampal mossy fiber-CA3 synapses. *Canadian journal of physiology and pharmacology* **95,** 1058–1063 (2017).

11. Moreira, I. S., Koukos, P. I., Melo, R., **Almeida, J. G.**, *et al.* SpotOn: high accuracy identification of protein-protein interface hot-spots. *Scientific reports* **7,** 1–11 (2017).

12. Sensoy, O., **Almeida, J. G.**, Shabbir, J., Moreira, I. S. & Morra, G. Computational studies of G protein-coupled receptor complexes: Structure and dynamics. *Methods in cell biology* **142,** 205–245 (2017).

Conference presentations

2021
The Natural History of Clonal Haematopoiesis, *CRUK Cambridge Centre Early Detection Programme 6th Annual Symposium*, Cambridge

2020
Leveraging Automated Blood Cell Morphology for Myelodysplastic Syndrome Diagnosis and Prognosis Prediction, *Quantitative BioImaging Conference*, Oxford

2017
Using big-data to understand the protein interface landscape, *Encontro de Jovens Investigadores de Biologia Computacional Estrutural*, Coimbra

2016
A Machine Learning Based Protein-Protein Hot-Spot Prediction Method - SpotOn, *Encontro de Jovens Investigadores de Biologia Computacional Estrutural*, Lisbon