





# José Guilherme de Almeida

*PhD, computational biologist (they/them)*

---

✉ jose.almeida@research.fchampalimaud.org     0000-0002-1887-0157  
 josegcpa.ml     github.com/josegcpa     bitbucket.org/josegcpa

## Education

---

Oct 2017 – 2022	<b>European Bioinformatics Institute and University of Cambridge</b> <i>PhD in Computational Biology</i>
Sep 2015 – Sep 2017	<b>Universidade de Coimbra</b> <i>MSc in Cell and Molecular Biology</i> with honours
Sep 2012 – Jul 2015	<b>Universidade de Coimbra</b> <i>BSc in Biochemistry</i>

## Relevant Research Experience

---

2022-current	<b>Computational Clinical Imaging Group @ Champalimaud Foundation</b> Under Nikolaos Papanikolaou <i>Post-doctoral fellow</i> <ul style="list-style-type: none"><li>▪ Development of robust machine- and deep-learning methods for prostate cancer detection and segmentation in multi-parametric magnetic resonance imaging</li><li>▪ Creation of reproducible pipelines for image registration and radiomic feature extraction</li></ul>
2017 - 2022	<b>Cancer data science group @ EMBL-EBI</b> Supervised by Moritz Gerstung and George S. Vassiliou <i>PhD fellow</i> <ul style="list-style-type: none"><li>▪ 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</li><li>▪ 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</li></ul>
2016 - 2017	<b>Data-driven molecular design group @ CNC-UC</b> Supervised by Irina S. Moreira <i>MSc student</i> <ul style="list-style-type: none"><li>▪ Development of machine-learning protocols to determine hot-spots (important residues) in the binding interfaces of proteins</li><li>▪ Structural and statistical analysis of large collections of protein-protein complexes and structural characterization of complexes with no known structure</li></ul>

## Professional certificates

---

2021	<b>Colorado Boulder University</b> <i>Statistical Inference for Data Science Applications Specialization (online)</i>
2021	<b>Erasmus University Rotterdam</b> <i>Econometrics course (online)</i>

## Technical skills

---

Programming	<b>Python</b> (advanced user), <b>R</b> (advanced user), <b>C</b> (beginner)
Machine-learning	<b>Machine-learning</b> — scikit-learn (Python), caret (R) <b>Deep-learning</b> — tensorflow, pytorch and related packages (Python)
Computer-vision	scikit-image, OpenCV (Python)
Statistical analysis	<b>Frequentist methods</b> <b>Bayesian methods</b> (particularly MCMC)
Data visualization	ggplot2 (R)
Workflows	<b>Containerisation</b> (Docker) <b>Workflow management</b> (snakemake)

## Teaching Experience

---

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

## Fellowships, awards and honors

---

2017	<b>Universidade de Coimbra</b> <i>Merit fellowship for exceptional curricular performance during my MSc</i>
2017 - 2022	<b>European Molecular Biology Laboratory</b> and support from the <b>National Health Research Institute</b> <i>PhD fellowship</i>

## Peer-reviewed research grants

---

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

## Other activities

---

2016 - 2017	<b>Junior Enterprise for Science and Technology (JEST)</b> <i>Co-founder</i> 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	<b>20th EMBL PhD Symposium</b> <i>Organization, speaker contact</i> Contacted different high-profile researchers to invite them to present at the 20th EMBL PhD Symposium
2019	<b>EBI-Sanger-Cambridge PhD Symposium (eSCAMPS) 2019</b> <i>Website design</i> Developed and designed the website for the 2019 eSCAMPS

## Languages

---

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

## Scientific publications

---

- |      |   |
|------|---|
| 2022 | 1. <b>De Almeida, J. G.</b> , Gudgin, E., Besser, M., Dunn, W. G., <i>et al.</i> Computational analysis of peripheral blood smears detects disease-associated cytomorphologies. <i>medRxiv</i> (2022).  |
|      | 2. Fabre, M. A., <b>de Almeida, J. G.</b> , Fiorillo, E., Mitchell, E., <i>et al.</i> The longitudinal dynamics and natural history of clonal haematopoiesis. <i>Nature (shared first-authorship with M Fabre)</i> , 1–8 (2022).  |
| 2021 | 3. Preto, A. J., Matos-Filipe, P., <b>de Almeida, J. G.</b> , Mourão, J. & Moreira, I. S. in <i>Artificial Neural Networks</i> 267–288 (Springer, 2021).  |
| 2020 | 4. Preto, A. J., Barreto, C. A., Baptista, S. J., Almeida, J. G. d., <i>et al.</i> Understanding the binding specificity of G-Protein coupled receptors toward G-proteins and arrestins: Application to the dopamine receptor family. <i>Journal of Chemical Information and Modeling</i> <b>60</b> , 3969–3984 (2020). |
| 2019 | 5. R Magalhães, P., Machuqueiro, M., G Almeida, J., Melo, A., <i>et al.</i> Dynamical rearrangement of human epidermal growth factor receptor 2 upon antibody binding: effects on the dimerization. <i>Biomolecules</i> <b>9</b> , 706 (2019).  |
| 2018 | 6. Lemos, A., Melo, R., Preto, A. J., <b>Almeida, J. G.</b> , <i>et al.</i> In silico studies targeting G-protein coupled receptors for drug research against Parkinson's disease. <i>Current neuropharmacology</i> <b>16</b> , 786–848 (2018).   |
|      | 7. Melo, R., Lemos, A., Preto, A. J., <b>Almeida, J. G.</b> , <i>et al.</i> Computational approaches in antibody-drug conjugate optimization for targeted cancer therapy. <i>Current topics in medicinal chemistry</i> <b>18</b> , 1091–1109 (2018).  |
|      | 8. Melo, R., Lemos, A., Preto, A. J., Bueschbell, B., <i>et al.</i> An Overview of Antiretroviral Agents for Treating HIV Infection in Paediatric Population. <i>Current medicinal chemistry</i> (2018).  |
|      | 9. Preto, A. J., <b>Almeida, J. G.</b> , Schaarschmidt, J., Xue, L. C., <i>et al.</i> Computational Tools for the Structural Characterization of Proteins and Their Complexes from Sequence-Evolutionary Data. <i>Encyclopedia of Analytical Chemistry: Applications, Theory and Instrumentation</i> , 1–19 (2018).     |
| 2017 | 10. <b>Almeida, J. G.</b> , Preto, A. J., Koukos, P. I., Bonvin, A. M. & Moreira, I. S. Membrane proteins structures: A review on computational modeling tools. <i>Biochimica et Biophysica Acta (BBA)-Biomembranes</i> <b>1859</b> , 2021–2039 (2017).   |
|      | 11. Bastos, F. C., Corceiro, V. N., Lopes, S. A., <b>de Almeida, J. G.</b> , <i>et al.</i> Effect of tolbutamide on tetraethylammonium-induced postsynaptic zinc signals at hippocampal mossy fiber-CA3 synapses. <i>Canadian Journal of Physiology and Pharmacology</i> <b>95</b> , 1058–1063 (2017).                  |
|      | 12. Moreira, I. S., Koukos, P. I., Melo, R., <b>Almeida, J. G.</b> , <i>et al.</i> SpotOn: high accuracy identification of protein-protein interface hot-spots. <i>Scientific reports</i> <b>7</b> , 1–11 (2017).   |
|      | 13. Sensoy, O., <b>Almeida, J. G.</b> , Shabbir, J., Moreira, I. S. & Morra, G. in <i>Methods in Cell Biology</i> 205–245 (Academic Press, 2017).   |

## Conference presentations

---

- |      |  |
|------|--|
| 2021 | The Natural History of Clonal Haematopoiesis, <i>CRUK Cambridge Centre Early Detection Programme 6th Annual Symposium</i> , Cambridge                                |
| 2020 | Leveraging Automated Blood Cell Morphology for Myelodysplastic Syndrome Diagnosis and Prognosis Prediction, <i>Quantitative Bioluminescence Conference</i> , Oxford  |
| 2017 | Using big-data to understand the protein interface landscape, <i>Encontro de Jovens Investigadores de Biologia Computacional Estrutural</i> , Coimbra                |
| 2016 | A Machine Learning Based Protein-Protein Hot-Spot Prediction Method — SpotOn, <i>Encontro de Jovens Investigadores de Biologia Computacional Estrutural</i> , Lisbon |