

Bernardo P. de Almeida

I am an AI Research Scientist interested in building large language foundational models for biology, particularly models that can read the human genome and interpret its variation.

Born 14 Dec 1994

Portuguese citizenship

E-mail: bernardo.almeida94@gmail.com

Website: <https://bernardo-de-almeida.github.io>

EDUCATION

2018 – 2023	PhD in Molecular Biosciences University of Vienna and Medical University of Vienna, Austria	[Passed with distinction]
2015 – 2017	Master in Oncobiology University of Algarve, Portugal	[Classification: 19/20 ; thesis: 20/20]
2012 – 2015	Bachelor in Biomedical Sciences University of Algarve, Portugal	[Class.: 17/20 ; thesis: 20/20]

RESEARCH EXPERIENCE

2023 –	Research Scientist in Artificial Intelligence and Biology InstaDeep, Paris, France Topic: “ <u>Development of large language foundational models for genomics</u> ”
2018 – 2023	PhD in Molecular Biosciences University of Vienna and Medical University of Vienna, Austria Research Institute of Molecular Pathology (IMP), Vienna, Austria Project: “ <u>Decoding the cis-regulatory information of enhancer sequences</u> ” Supervisor: Dr Alexander Stark
2016 – 2018	Master’s thesis + Researcher Instituto de Medicina Molecular, Lisboa, Portugal Project: “Discovery of novel mechanisms of centrosome amplification and their therapeutic value in cancer” Supervisor: Dr Nuno Barbosa Morais (collab. Dr Mónica Bettencourt Dias)
2016	Visiting Worker Gaffney Group, <u>Wellcome Trust Sanger Institute, Cambridge, UK</u> Project: “Map of histone Quantitative Trait Loci (QTLs) in iPSCs” Supervisor: Dr Ângela Gonçalves & Dr Daniel Gaffney
2015 – 2016	BSc’s final project + Research fellow Centre for Biomedical Research, University of Algarve, Portugal Projects: “Identification of new <u>genetic risk markers</u> for breast cancer” & “ <u>Cis-regulation of somatic mutations in breast and ovarian cancers</u> ” Supervisor: Professor Ana Teresa Maia

AWARDS

2024	2024 Denise P. Barlow Award for best PhD thesis on biological mechanisms Vienna, Austria
2023	2023 Vienna BioCenter PhD Award Vienna BioCenter, Austria
2022	Life Science Research Award Austria 2022 - category Basic Science Austrian Society for Molecular Biosciences and Biotechnology (ÖGMBT)
2018	Best Master Student of the Sciences and Technologies field University of Algarve, Portugal

Selected:

25. **B.P. de Almeida**, C. Schaub, M. Pagani, S. Secchia, E. Furlong, A. Stark. "Targeted design of synthetic enhancers for selected tissues in the *Drosophila* embryo". **Nature** 2024
. *Featured commentary*: Alexandra Despang. "Designer enhancers for cell-type-specific gene regulation". *Nature Biotechnology* 2024
24. F. Reiter*, **B.P. de Almeida***, A. Stark. "Enhancers display constrained sequence flexibility and context-specific modulation of motif function". **Genome Research** 2023; 33:346-358
23. **B.P. de Almeida**, F. Reiter, M. Pagani, A. Stark. "DeepSTARR predicts enhancer activity from DNA sequence and enables the *de novo* design of synthetic enhancers". **Nature Genetics** 2022; 54:613-624
. *Featured commentary*: Lin Tang. "Predicting and designing enhancers". *Nature Methods* 2022
. Awarded the Life Science Research Award Austria 2022, by ÖGMBT
22. **B.P. de Almeida**, A.F. Vieira, J. Paredes, M. Bettencourt-Dias, N.L. Barbosa-Morais. "Pan-cancer association of a centrosome amplification gene expression signature with genomic alterations and clinical outcome". **PLoS Computational Biology** 2019; 15(3):e1006832
21. **B.P. de Almeida***, J.D. Apolonio*, A. Binnie, P. Castelo-Branco. "Roadmap of DNA methylation in breast cancer identifies novel prognostic biomarkers". **BMC Cancer** 2019

Preprints:

20. G. Richard*, **B.P. de Almeida***, H. Dalla-Torre, C. Blum, L. Hexemer, P. Pandey, S. Laurent, M. Lopez, A. Laterre, M. Lang, U. Şahin, K. Beguir, T. Pierrot. "ChatNT: A Multimodal Conversational Agent for DNA, RNA and Protein Tasks". **bioRxiv** 2024
19. **B.P. de Almeida***, H. Dalla-Torre*, G. Richard, C. Blum, L. Hexemer, M. Gélard, P. Pandey, S. Laurent, A. Laterre, M. Lang, U. Şahin, K. Beguir, T. Pierrot. "SegmentNT: annotating the genome at single-nucleotide resolution with DNA foundation models". **bioRxiv** 2024
18. E. Trop*, C.-H. Kao*, M. Polen*, Y. Schiff*, **B.P. de Almeida**, A. Gokaslan, T. Pierrot, V. Kuleshov. "Advancing DNA Language Models: The Genomics Long-Range Benchmark". **MLGenX ICLR Workshop** 2024
17. J. J. Garau-Luis*, P. Bordes*, L. Gonzalez*, M. Roller, **B. P. de Almeida**, L. Hexemer, C. Blum, S. Laurent, J. Grzegorzewski, M. Lang, T. Pierrot§, G. Richard§. "Multi-modal Transfer Learning between Biological Foundation Models". **arXiv** 2024
16. V. Loubiere, **B.P. de Almeida**, M. Pagani, A. Stark. "Developmental and housekeeping transcriptional programs display distinct modes of enhancer-enhancer cooperativity in *Drosophila*". **bioRxiv** 2023
15. N. Moreno-Marin, G. Marteil, N.C. Fresmann, **B.P. de Almeida**, ..., J.B. Pereira-Leal, J.T. Barata, S. Godinho, N.L. Barbosa-Morais, M. Bettencourt-Dias. "High prevalence and dependence of centrosome clustering in mesenchymal tumors and leukemia". **bioRxiv** 2023
14. J.M. Xavier, R. Magno, R. Russell, **B.P. de Almeida**, A. Jacinta-Fernandes, A. Duarte, M. Dunning, S. Samarajiwa, M. O'Reilly, C.L. Rocha, N. Rosli, B.A.J. Ponder, A.T. Maia. "Mapping of cis-regulatory variants by differential allelic expression analysis identifies candidate risk variants and target genes of 27 breast cancer risk loci". **medRxiv** 2022

Others:

13. S. Boshar, E. Trop, **B.P. de Almeida**, L. Copoiu, T. Pierrot. "Are genomic language models all you need? Exploring genomic language models on protein downstream tasks". **Bioinformatics** 2024
12. H. Dalla-Torre, L. Gonzalez, J. Mendoza-Revilla, N.L. Carranza, A.H. Grzywaczewski, F. Oteri, C. Dallago, E. Trop, **B.P. de Almeida**, H. Sirelkhatim, G. Richard, M. Skwark, K. Beguir, M. Lopez[§], T. Pierrot[§]. "The Nucleotide Transformer: Building and Evaluating Robust Foundation Models for Human Genomics". **Nature Methods** 2024
11. J. Mendoza-Revilla, E. Trop, L. Gonzalez, M. Roller, H. Dalla-Torre, **B.P. de Almeida**, G. Richard, J. Caton, N.L. Carranza, M. Skwark, A. Laterre, K. Beguir, T. Pierrot[§], M. Lopez[§]. "A Foundational Large Language Model for Edible Plant Genomes". **Commun Biol** 2024
10. L. Klaus, **B.P. de Almeida**, A. Vlasova, F. Nemčko, A. Schleiffer, K. Bergauer, L. Hofbauer, M. Rath, A. Stark. "Identification and characterization of repressive domains in Drosophila transcription factors". **The EMBO Journal** 2022, e112100
9. L. Correia, R. Magno, J.M. Xavier, **B.P. de Almeida**, F. Esteves, I. Duarte, M. Eldridge, C. Sun, A. Bosma, L. Mittempergher, A. Marreiros, R. Bernards, C. Caldas, S.F. Chin[§], A.T. Maia[§]. "Allelic expression imbalance of PIK3CA mutations is frequent in breast cancer and prognostically significant". **npj Breast Cancer** 2022; 8:71
8. J. Conde*, R.A. Pumroy*, C. Baker*, T. Rodrigues*, A. Guerreiro, B.B. Sousa, M.C. Marques, **B.P. de Almeida**, ... , V.Y. Moiseenkova-Bell[§], G.J.L. Bernardes[§]. "Allosteric Antagonist Modulation of TRPV2 by Piperlongumine Impairs Glioblastoma Progression". **ACS Central Science** 2021; 7(5):868-881
7. I. Gomes, **B.P. de Almeida**, S. Dâmaso, A. Mansinho, I. Correia, S. Henriques, R. Cruz-Duarte, G. Vilhais, P. Félix, P. Alves, P. Corredeira, N.L. Barbosa-Morais, L. Costa, S. Casimiro. "Expression of receptor activator of NFkB (RANK) drives stemness and resistance to therapy in ER+HER2- breast cancer". **Oncotarget** 2020; 11(19):1714-1728
6. T. Rodrigues, **B.P. de Almeida**, N.L. Barbosa-Morais, G.J.L. Bernardes. "Dissecting celastrol with machine learning to unveil dark pharmacology". **Chemical Communications** 2019; 55:6369-6372
5. C. Baker, T. Rodrigues, **B.P. de Almeida**, N.L. Barbosa-Morais, G.J.L. Bernardes. "Natural product-drug conjugates for modulation of TRPV1-expressing tumors". **Bioorganic & Medicinal Chemistry** 2019; 27(12):2531-2536
4. S. Braun, M. Enculescu, S.T. Setty, M. Cortés-López, **B.P. de Almeida**, F.X.R. Sutandy, L. Schulz, A. Busch, M. Seiler, S. Ebersberger, N.L. Barbosa-Morais, S. Legewie, J. König, K. Zarnack. "Decoding a cancer-relevant splicing decision in the RON proto-oncogene using high-throughput mutagenesis". **Nature Communications** 2018; 9:3315
3. G. Marteil, A. Guerrero, A.F. Vieira, **B.P. de Almeida**, P. Machado, S. Mendonça, M. Mesquita, B. Villarreal, I. Fonseca, M.E. Francia, K. Dores, N.P. Martins, S.S. Jana, E. Tranfield, N.L. Barbosa-Morais, J. Paredes, D. Pellman, S.A. Godinho, M. Bettencourt-Dias. "Over-elongation of centrioles in cancer promotes centriole amplification and chromosome missegregation". **Nature Communications** 2018; 9:1258

Workshop papers:

2. E. Trop, C.-H. Kao, M. Polen, Y. Schiff, **B. P. de Almeida**, A. Gokaslan, T. Pierrot, V. Kuleshov. "Advancing DNA language models: The genomics long-range benchmark" **LLMs4Bio AAAI Workshop 2024, MLGenX ICLR Workshop 2024**
1. S. Boshar, E. Trop, **B.P. de Almeida**, T. Pierrot. "Exploring Genomic Language Models on Protein Downstream Tasks". **LLMs4Bio AAAI Workshop 2024, MLGenX ICLR Workshop 2024**

SOFTWARES

B.P. de Almeida*, N. Saraiva-Agostinho*, N.L. Barbosa-Morais. "cTRAP: Identification of candidate causal perturbations from differential gene expression data". **R package**, <https://bioconductor.org/packages/release/bioc/html/cTRAP.html>

PATENTS

"Methods of cancer treatment". Intellectual Patent Office UK, Provisional Patent Application GB 1820975.9. Authors: N.L. Barbosa-Morais, **B.P. de Almeida**, M. Bettencourt-Dias, J. Paredes, A. Vieira (2018).

CO-SUPERVISION

Monika Heinzl, PhD student

. Research Institute of Molecular Pathology (IMP), Vienna, Austria *Oct 2022 – May 2023*

Luís Bento, 6-month internship (Master's student in Biological Engineering)

. Instituto de Medicina Molecular, Lisbon, Portugal *Mar – Sept 2017*

TEACHING EXPERIENCE

Decoding the genome with deep learning

. OLISPO Workshop, INESC-ID, Lisbon, Portugal *27 Apr 2023*

INVITED TALKS

DeepSTARR predicts enhancer activity from DNA sequence and enables the de novo design of enhancers

. Kipoi seminar series (<https://tinyurl.com/bd39r2yz>) *5 Apr 2023*

. MLCSB - ISCBacademy Webinar (<https://tinyurl.com/53drm9k5>) *9 Dec 2021*

Decoding transcriptional regulation using massively parallel reporter assays and Twist Oligo Pools

. Twist Bioscience Webinar, Virtual (<https://tinyurl.com/3nuw7wsw>) *14 Apr 2021*

PARTICIPATION IN MEETINGS

(*equal contributions, §co-corresponding authors)

Oral presentations:

The Nucleotide Transformer: building and evaluating robust foundation models for genomics

B.P. de Almeida, G. Richard, H. Dalla-Torre, J. Mendoza-Revilla, M. Lopez, A. Laterre, K. Beguir, T. Pierrot

. Systems Biology: Global Regulation of Gene Expression, CSHL, USA *13-16 Mar 2024*

. Machine Learning and Computational Biology 2024 *5-6 Sept 2024*

DeepSTARR predicts enhancer activity from DNA sequence and enables the *de novo* design of enhancers

B.P. de Almeida, F. Reiter, M. Pagani, A. Stark

. Systems Biology: Global Regulation of Gene Expression, CSHL, USA 9-12 Mar 2022

. EMBO Workshop Enhanceropathies: Understanding enhancer function to understand human disease 6-9 Oct 2021

Discovery of novel mechanisms of centrosome amplification and their therapeutic value in cancer

B.P. de Almeida, G. Marteil, M. Bettencourt-Dias, N.L. Barbosa-Morais

. iMed Conference® 9.0, Lisbon, Portugal (2nd *place at Innovate Competition*) 25-29 Oct 2017

. 12th Young European Scientists meeting, Porto, Portugal 14-17 Sept 2017

Poster Presenter:

Enhancers display constrained sequence flexibility and context-specific modulation of motif function

B.P. de Almeida*, F. Reiter*, A. Stark

. 15th EMBL Conference: Transcription and Chromatin, Heidelberg, Germany 27-30 Aug 2022

Understanding the contribution of inter-motif spacer sequences to enhancer activity

B.P. de Almeida, F. Reiter, A. Stark

. 11th Visualizing Biological Data meeting (VIZBI), *Virtual* 24-26 Mar 2021

. 14th EMBL Conference: Transcription and Chromatin, *Virtual* 27-29 Aug 2020

Pan-cancer analysis of Centrosome Amplification uncovers its association with copy number alterations and poor clinical outcome (*highlighted poster*)

B.P. de Almeida, N.L. Barbosa-Morais

. 3rd ASPIC International Congress, Lisbon, Portugal 10-11 May 2018

Discovery of novel mechanisms of centrosome amplification and their therapeutic value in cancer

B.P. de Almeida, G. Marteil, A. Guerrero, M. Bettencourt-Dias, N.L. Barbosa-Morais

. 3rd EACR Conference in Cancer Genomics, Cambridge, UK 25-28 June 2017

PIK3CA mutant allele differential expression (MADE) association analysis with breast cancer

B.P. de Almeida, J.M. Xavier, C. Sun, I.A. Silva, J.J. Silva, A. Marreiros, M. Eldridge, R. Bernards, C. Caldas, S.F. Chin, A.T. Maia

. 2nd ASPIC International Congress, Porto, Portugal 28-29 Apr 2016

Poster Abstracts (presenter's name underlined):

SegmentNT: annotating the genome at single-nucleotide resolution with DNA foundation models

B.P. de Almeida*, H. Dalla-Torre*, G. Richard, C. Blum, L. Hexemer, M. Gélard, P. Pandey, S. Laurent, A. Laterre, M. Lang, U. Şahin, K. Beguir, T. Pierrot

. Systems Biology: Global Regulation of Gene Expression, CSHL, USA 13-16 Mar 2024

Distinct enhancer-enhancer cooperative behaviours underlie developmental and housekeeping transcription in *Drosophila*

V. Loubiere, **B.P. de Almeida**, M. Pagani, A. Stark

. Systems Biology: Global Regulation of Gene Expression, CSHL, USA 13-16 Mar 2024

. 15th EMBL Conference: Transcription and Chromatin, Heidelberg, Germany 27-30 Aug 2022

Identification of repressive protein domains and their interacting co-repressors

L. Klaus, A. Vlasova, **B.P. de Almeida**, F. Nemcko, A. Schleiffer, K. Bergauer, M. Rath, A. Stark

. 15th EMBL Conference: Transcription and Chromatin, Heidelberg, Germany 27-30 Aug 2022

Transcriptional enhancer activity relies on specific TF motif compatibilities (*poster prize*)

F. Reiter*, **B.P. de Almeida***, A. Stark

. EMBO Workshop Enhanceropathies: Understanding enhancer function to understand human disease 6-9 Oct 2021

cTRAP: identification of candidate causal perturbations from differential expression data

N. Saraiva-Agostinho, **B.P. de Almeida**, N.L. Barbosa-Morais

. 11th Visualizing Biological Data meeting (VIZBI), *Virtual* 24-26 Mar 2021

Characterization of enhancer-bound proteomes

F. Reiter, **B.P. de Almeida**, R. Imre, K. Mechtler, A. Stark

. 14th EMBL Conference: Transcription and Chromatin, *Virtual* 27-29 Aug 2020

Biological features of estrogen receptor-positive breast cancer with elevated RANK (TNFRSF11A) expression

S. Casimiro, I. Gomes, **B.P. de Almeida**, P. Alves, P. Félix, G. Vilhais, A. Mansinho, M.R. Dionísio, N.L. Barbosa-Morais, L. Costa

. 2019 ASCO Annual Meeting, Chicago, USA 31 May – 04 June 2019

Integrative genomic approach elucidates the risk mechanism for breast cancer associated 5q14.1 locus (*highlighted poster*)

J.M. Xavier, R. Magno, **B.P. de Almeida**, M. Dunning, A. Jacinta-Fernandes, R. Russell, S. Samarajiwa, M. O'Reilly, N. Rosli, C. Nobrega, N.L. Barbosa-Morais, C. Caldas, B.A. Ponder, A.T. Maia

. 3rd ASPIC International Congress, Lisbon, Portugal 10-11 May 2018

Mapping of cis-regulatory variants helps dissecting the risk mechanism for breast cancer associated 5q14.1 locus

J.M. Xavier, R. Magno, **B.P. de Almeida**, M. Dunning, A. Jacinta-Fernandes, R. Russell, S. Samarajiwa, M. O'Reilly, N. Rosli, C. Nobrega, N.L. Barbosa-Morais, C. Caldas, B.A.J. Ponder, A.T. Maia

. 21^a Reunião da Sociedade Portuguesa de Genética Humana, Portugal 16-18 Nov 2017

Roadmap of DNA methylation in breast cancer identifies 15 novel potential biomarkers

B.P. de Almeida, J.D. Apolonio, A. Binnie, P. Castelo-Branco

. 2nd CBMR/ProRegem Annual Meeting, University of Algarve, Portugal 8-9 Sept 2017

Analysis of potential cis-regulatory variants at locus 17q22

F. Esteves, J. Xavier, R. Magno, **B.P. de Almeida**, A. Fernandes, C. Rocha, A.T. Maia

. 2nd CBMR/ProRegem Annual Meeting, University of Algarve, Portugal 8-9 Sept 2017

PIK3CA mutant allele differential expression (MADE) associates with breast cancer clinical features

J.M. Xavier, **B.P. de Almeida**, C. Sun, J. Silva, A. Marreiros, M. Eldridge, R. Bernards, C. Carlos, S.F. Chin, A.T. Maia

. 24th Biennial EACR Congress, Manchester, UK 9-12 July 2016

Integrative differential allelic expression analysis efficiently reveals the biology underlying risk to breast cancer

J.M. Xavier, R. Russell, **B.P. de Almeida**, N. Rosli, C. Rocha, S. Samarajiwa, S.F. Chin, C. Caldas, B.A.J. Ponder, A.T. Maia

. 2nd ASPIC International Congress, Porto, Portugal 28-29 April 2016

. AACR Conference on Advances in Breast Cancer Research, Washington, USA 17-20 Oct 2015

LANGUAGES

Portuguese (native), English (fluent), Spanish (fluent), French (basics), German (basics)