

# PEDRO L. BALDONI

pedrobaldoni [at] gmail.com

<https://plbaldoni.rbind.io>

April 9, 2021

## EDUCATION

---

**Ph.D., Biostatistics** 2020

Department of Biostatistics

University of North Carolina at Chapel Hill, United States

Advisors: Dr. Naim U. Rashid, Dr. Joseph G. Ibrahim

Committee: Dr. Michael I. Love, Dr. Yun Li, and Dr. Douglas Phanstiel

Dissertation title: *Statistical Methods for the Analysis of Epigenomic Data*

**M.S., Statistics** 2014

Department of Statistics

University of Campinas, Brazil

Advisor: Dr. Hildete P. Pinheiro

International exchange program: Aarhus University, Denmark

Thesis title: *Multivariate Generalized Linear Mixed Models for Genetic Characterization of Diseases*

**B.S., Statistics** 2011

Department of Statistics

University of Campinas, Brazil

## PROFESSIONAL EXPERIENCE

---

**Postdoctoral Fellow** 2020 - present

Bioinformatics Division

Walter and Eliza Hall Institute of Medical Research, Australia

Supervisors: Professor Gordon K. Smyth and Associate Professor Andrew I. Webb

**Graduate Research Assistant** 2015 - 2020

Department of Biostatistics

University of North Carolina at Chapel Hill, United States

Supervisors: Dr. Michael G. Hudgens, Dr. Jianwen Cai, and Dr. Joseph G. Ibrahim

**Statistician** 2012

CPqD Foundation, Brazil

## HONORS, AWARDS AND SCHOLARSHIPS

---

**University Cancer Research Fund Award** 2019 - 2020

Lineberger Comprehensive Cancer Center

University of North Carolina at Chapel Hill, United States

<b>Max Halperin Award (Excellence in Graduate Studies)</b> Department of Biostatistics University of North Carolina at Chapel Hill, United States	2016
<b>Science Without Borders (Ph.D. program)</b> Coordination for the Improvement of Higher Education Personnel Ministry of Education, Brazil	2014 - 2018
<b>International Exchange Program at Aarhus University, Denmark</b> Santander Bank, Brazil	2013 - 2014
<b>Scholarship (M.S. program)</b> Coordination for the Improvement of Higher Education Personnel Ministry of Education, Brazil	2012 - 2014

## PROFESSIONAL AND EDITORIAL ACTIVITIES

---

### Professional Memberships

Australian Bioinformatics and Computational Biology Society (ABACBS)	2020 - present
American Statistical Association (ASA)	2018 - present
Eastern North American Region (ENAR)	2017 - present

### Peer Review Activities

Journal Referee: Nucleic Acids Research

### Program Development

Session Chair, ENAR Conference. Topic: Replicability in Big Data Precision Medicine. 2019

## PUBLICATIONS

---

\* indicates equal contributions

† indicates corresponding author(s) (if not the senior author)

### Journal Articles (peer reviewed)

- [0] **P. L. Baldoni**<sup>†</sup>, N. U. Rashid, and J. G. Ibrahim. “Efficient Detection and Classification of Epigenomic Changes Under Multiple Conditions”. In: *Biometrics (in press)* (2021).
- [1] **P. L. Baldoni**, D. Sotres-Alvarez, T. Lumley, and P. A. Shaw<sup>†</sup>. “On the use of Regression Calibration in a Complex Survey Design with Application to the Hispanic Community Health Study/Study of Latinos”. In: *American Journal of Epidemiology* (2021). DOI: [10.1093/aje/kwab008](https://doi.org/10.1093/aje/kwab008).
- [2] T. Elfassy<sup>†</sup>, A. Zeki Al Hazzouri, J. Cai, **P. L. Baldoni**, M. M. Llabre, T. Rundek, L. Raij, J. P. Lash, G. A. Talavera, S. Wassertheil-Smoller, M. Daviglus, J. N. Booth III, S. F. Castaneda, M. Garcia, and N. Schneiderman. “Incidence of Hypertension Among US Hispanics/Latinos: The Hispanic Community Health Study/Study of Latinos, 2008 to 2017”. In: *Journal of the American Heart Association* 9.12 (2020), pp. 1–13. DOI: [10.1161/JAHA.119.015031](https://doi.org/10.1161/JAHA.119.015031).

- [3] R. C. Kaplan<sup>†</sup>, **P. L. Baldoni**, G. M. Strizich, E. J. Pérez-Stable, N. L. Saccone, C. A. Peralta, K. M. Perreira, M. D. Gellman, J. S. Williams-Nguyen, C. J. Rodriguez, D. J. Lee, M. Daviglus, G. A. Talavera, J. P. Lash, J. Cai, and N. Franceschini. “Current smoking raises risk of incident hypertension: Hispanic Community Health Study-Study of Latinos”. In: *American Journal of Hypertension (online ahead of print)* (2020). DOI: [10.1093/ajh/hpaa152](https://doi.org/10.1093/ajh/hpaa152).
- [4] K. R. Mollan<sup>†</sup>, I. M. Trumble, S. A. Reifeis, O. Ferrer, C. P. Bay, **P. L. Baldoni**, and M. G. Hudgens. “Precise and accurate power of the rank-sum test for a continuous outcome”. In: *Journal of Biopharmaceutical Statistics* 30.4 (2020), pp. 639–648. DOI: [10.1080/10543406.2020.1730866](https://doi.org/10.1080/10543406.2020.1730866).
- [5] **P. L. Baldoni**<sup>†</sup>, N. U. Rashid, and J. G. Ibrahim. “Improved detection of epigenomic marks with mixed-effects hidden Markov models”. In: *Biometrics* 75.4 (2019), pp. 1401–1413. DOI: [10.1111/biom.13083](https://doi.org/10.1111/biom.13083).
- [6] S.-K. Lee<sup>†</sup>, S. Zhou, **P. L. Baldoni**, E. Spielvogel, N. M. Archin, M. G. Hudgens, D. M. Margolis, and R. Swanstrom<sup>†</sup>. “Quantification of the latent HIV-1 reservoir using ultra deep sequencing and primer ID in a viral outgrowth assay”. In: *Journal of Acquired Immune Deficiency Syndromes* 74.2 (2017), pp. 221–228. DOI: [10.1097/QAI.0000000000001187](https://doi.org/10.1097/QAI.0000000000001187).
- [7] I. M. Trumble<sup>†</sup>, A. G. Allmon, N. M. Archin, J. Rigdon, O. Francis, **P. L. Baldoni**, and M. G. Hudgens. “SLDAssay: A software package and web tool for analyzing limiting dilution assays”. In: *Journal of Immunological Methods* 450 (2017), pp. 10–16. DOI: [10.1016/j.jim.2017.07.004](https://doi.org/10.1016/j.jim.2017.07.004).
- [8] G. Clutton, Y. Xu, **P. L. Baldoni**, K. R. Mollan, J. Kirchherr, W. Newhard, K. Cox, J. D. Kuruc, A. Kashuba, R. Barnard, N. Archin, C. L. Gay, M. G. Hudgens, D. M. Margolis, and N. Goonetilleke<sup>†</sup>. “The differential short-and long-term effects of HIV-1 latency-reversing agents on T cell function”. In: *Scientific reports* 6.30749 (2016), pp. 1–16. DOI: [10.1038/srep30749](https://doi.org/10.1038/srep30749).

#### Preprints (not peer reviewed)

#### Preprints, subsequently published (not peer reviewed)

- [9] **P. L. Baldoni**<sup>†</sup>, N. U. Rashid, and J. G. Ibrahim. “Efficient Detection and Classification of Epigenomic Changes Under Multiple Conditions”. In: *bioRxiv (accepted for publication in Biometrics)* (2020), pp. 1–34. DOI: [10.1101/864124](https://doi.org/10.1101/864124).
- [10] K. R. Mollan<sup>†</sup>, I. M. Trumble, S. A. Reifeis, O. Ferrer, C. P. Bay, **P. L. Baldoni**, and M. G. Hudgens. “Exact power of the rank-sum test for a continuous variable”. In: *arXiv (published in Journal of Biopharmaceutical Statistics)* (2019), pp. 1–15. URL: <https://arxiv.org/abs/1901.04597v1>.

## PRESENTATIONS & SCIENTIFIC MEETINGS

### Contributed Talks

*On the Use of Regression Calibration in a Complex Sampling Design With Application to the Hispanic Community Health Study/Study of Latinos*, Joint Statistical Meetings (JSM), August 2021

*Efficient Detection and Classification of Epigenomic Changes Under Multiple Conditions*, East-

ern North American Region (ENAR), March 2020

*Detection and Classification of Changes in Protein-DNA Binding Activity With Applications in Diffuse ChIP-seq Data*, Joint Statistical Meetings (JSM), July 2019

*Integrative HMM With Mixture Model for Differential Pattern Detection of Broad Epigenomic Marks*, Eastern North American Region (ENAR), March 2019

*Improved Detection of Epigenomic Marks With Mixed Effects Hidden Markov Models*, Eastern North American Region (ENAR), March 2018

*A Statistical Method for the Analysis of Multiple ChIP-seq Datasets*, Eastern North American Region (ENAR), March 2017

### **Invited Talks**

*Title to be advised*, WEHI Bioinformatics Seminar, Walter and Eliza Hall Institute of Medical Research, November 2021

*Detecting and Classifying Epigenomic Changes Under Multiple Conditions and Single Cell Applications*, WEHI Bioinformatics Seminar, Walter and Eliza Hall Institute of Medical Research, September 2020

*Statistical Methods for ChIP-seq Data*, Laboratory for Statistical and Translational Genomics, University of Pennsylvania, January 2020

### **Posters**

*epigraHMM - Epigenomic r-based analysis with hidden Markov models*, Bioconductor Conference (BioC), July 2020

*A Statistical Method for the Analysis of Multiple ChIP-seq Datasets*, Joint Statistical Meetings (JSM), July 2017

### **Other Meetings and Events**

*Efficient Detection of Epigenomic Changes Under Multiple Conditions and Single-cell Applications*, School of Medicine, University of North Carolina at Chapel Hill, United States, April 2020

*On the use of Regression Calibration in a Complex Survey Design with Application to the Hispanic Community Health Study/Study of Latinos*, Collaborative Studies Coordinating Center, University of North Carolina at Chapel Hill, United States, November 2019

*Statistical Strategies for the Analysis of Diet-Disease Models that Correct for Error-Prone Exposures*, Collaborative Studies Coordinating Center, University of North Carolina at Chapel Hill, United States, May 2019

*Calculating and Comparing Age Standardized Cumulative Incidence of Hypertension across Hispanic/Latino Background Groups*, Collaborative Studies Coordinating Center, University of North Carolina at Chapel Hill, United States, April 2019

*Statistical methods for HIV-1 reservoir estimation in viral outgrowth assays*, Center for AIDS Research (CFAR), University of North Carolina at Chapel Hill, United States, June 2016

## TEACHING

---

### **Recitation Lecturer**

Department of Biostatistics, University of North Carolina at Chapel Hill, United States  
BIOS 545 - Principles of Experimental Analysis, 2016

### **Teaching Assistant**

Department of Biostatistics, University of North Carolina at Chapel Hill, United States  
BIOS 680 - Introductory Survivorship Analysis, 2018  
BIOS 735 - Introduction to Statistical Computing, 2019 - 2020  
  
Department of Statistics, University of Campinas, Brazil  
ME607 - Time Series, 2013

## SOFTWARE

---

### **epigraHMM**

<https://github.com/plbaldoni/epigraHMM>

R package for comprehensive analysis of epigenomic data that include pre- and post-processing, consensus and differential peak detection, and genomic segmentation

### **mixNBHMM**

<https://github.com/plbaldoni/mixNBHMM>

Detection and classification of differential enrichment regions from ChIP-seq experiments under multiple conditions

### **ZIMHMM**

<https://github.com/plbaldoni/ZIMHMM>

Detection of broad enrichment regions from multiple ChIP-seq experimental replicates via a zero inflated mixed effects hidden Markov model

## SKILLS AND SERVICE

---

Computing skills: proficient in R, Python, SAS, and C++.  
Languages: Portuguese (native speaker), English, and Italian.