

# Curriculum Vitae

Notarization. I have read the following and certify that this curriculum vitae is a current and accurate statement of my professional record.

---

Signature

---

Date

## **I Personal Information**

### **I.A UID, Last Name, First Name, Middle Name, Contact Information**

Hector Corrada Bravo  
3114F Biomolecular Sciences Building #296  
University of Maryland  
College Park, MD 20745  
hcorrada@umiacs.umd.edu  
<http://www.cbcb.umd.edu/~hcorrada>  
UID: 111544240

### **I.B Academic Appointments at UMD**

- Assistant Professor, Department of Computer Science  
University of Maryland, College Park, MD  
July 2010-Present
- Assistant Professor, Institute for Advanced Computer Studies  
University of Maryland, College Park, MD  
July 2010-Present

### **I.D Other Employment**

- Postdoctoral Fellow, Department of Biostatistics  
Johns Hopkins University School of Public Health, Baltimore, MD  
September 2008-June 2010
- Research and Teaching Assistant, Departments of Computer Science and Statistics  
University of Wisconsin, Madison, WI  
September 2003-August 2008
- Research Intern,  
IBM Research, Almaden, CA  
May 2005-August 2005

## I.E Educational Background

- Ph.D., Computer Science, University of Wisconsin, Madison, WI  
September 2003 - August 2008  
Dissertation: Graph-based data analysis  
Advisor: Grace Wahba and Raghu Ramakrishnan
- D.M.A., Indiana University School of Music, Bloomington, IN (ABD)  
September 2000 - August 2003
- M.M., Peabody Institute of Music, Baltimore, MD  
September 1997 - May 1999
- B.M., Peabody Institute of Music, Baltimore, MD  
September 1993 - May 1997

## II Research, Scholarly and Creative Activities

### II.C Articles in Refereed Journals

1. H.S. Parker<sup>^</sup>, **H. Corrada Bravo**, J.T. Leek<sup>\*</sup> (2014). Removing batch effects for prediction problems with frozen surrogate variable analysis. *PeerJ*, *in press*.
2. W. Timp<sup>^</sup>, **H. Corrada Bravo**<sup>^</sup>, O.G. McDonald, M. Goggins, C. Umbricht, M. Zeiger, A.P. Feinberg<sup>\*</sup>, R.A. Irizarry<sup>\*</sup> (2014). Large hypomethylated blocks related to large heterochromatin regions as a universal defining epigenetic alteration in human solid tumors. *Genome Medicine* 6 (61), doi:10.1186/s13073-014-0061-y.
3. F. Chelaru<sup>^#</sup>, L. Smith<sup>#</sup>, N. Goldstein<sup>#</sup>, **H. Corrada Bravo**<sup>\*</sup> (2014). Epiviz: interactive visual analytics for epigenomics data. *Nature Methods*, doi:10.1038/nmeth.3038.
4. Pop, M.<sup>^</sup>, Walker, A.W., Paulson, J.<sup>#</sup>, Lindsey, B., Antonio, M., Hossain, M.A., Oundo, J., Tamboura, B., Mai, V., Astrovskaya, I., **Corrada Bravo**, **H.**, Rance, R., Stares, M., Levine, M.M., Panchalingam, S., Kotloff, K., Ikumapayi, U.N., Ebruke, C., Adeyemi, D., Ahmed, F., Alam, M.T., Amin, R., Siddiqui, S., Ochieng, J.B., Ouma, E., Juma, J., Mailu, E., Omore, R., Morris, J.G., Breiman, R.F., Saha, D., Parkhill, J., Stine, O.C.<sup>\*</sup>, Nataro, J.P.<sup>\*</sup> (2014). Diarrhea in young children from low-income countries leads to large-scale alterations in intestinal microbiota composition. *Genome Biology* 15, R76.
5. M. Aryee<sup>^</sup>, A. Jaffe, **H. Corrada Bravo**, C. Ladd-Acosta, A. Feinberg, K. Hansen<sup>\*</sup>, R.A. Irizarry<sup>\*</sup> (2014). Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA Methylation microarrays. *Bioinformatics* 30 (10):1363-9.
6. C. Ye<sup>^#</sup>, C. Hsiao<sup>#</sup>, **H. Corrada Bravo**<sup>\*</sup> (2014). BlindCall: ultra-fast base-calling of second-generation sequencing by blind deconvolution. *Bioinformatics* 30 (9):1214-9.
7. E. Alemu<sup>^</sup>, J.W. Carl, **H. Corrada Bravo**<sup>\*</sup>, S. Hannenhalli<sup>\*</sup> (2014). Determinants of expression variability. *Nucleic Acids Research* 42 (6), 3503-14.
8. N. Akula<sup>^</sup>, J. Barb, X. Jiang, J. Wendland, K. Choi, S. Sen, L. Hou, D. Chen, G. Laje, K. Johnson, B. Lipska, J. Kleinman, **H. Corrada Bravo**, S. Detera-Wadleigh, P.J. Munson, F.J. McMahon<sup>\*</sup> (2014). RNA-sequencing of brain transcriptome implicates dysregulation of neuroplasticity, circadian rhythms and GTPase binding in bipolar disorder. *Molecular Psychiatry*, doi:10.1038/mp.2013.170.

9. J. Paulson<sup>#</sup>, O.C. Stein, **H. Corrada Bravo**<sup>\*</sup>, M. Pop<sup>\*</sup> (2013). Differential abundance analysis for microbial marker-gene surveys. *Nature Methods* 10 (12):1200-1202, doi:10.1038/nmeth.2658.
10. X. He<sup>^</sup>, R. Chatterjee<sup>^</sup>, S. John, **H. Corrada Bravo**, B.K. Sathyanarayana, S.C. Biddle, P.C. Fitzgerald, J.A. Stamatoyannopoulos, G.L. Hager, C. Vinson<sup>\*</sup> (2013). Contribution of nucleosome binding preferences and co-occurring DNA sequences to transcription factor binding. *BMC Genomics* 14 (428).
11. S. Boca<sup>^</sup>, **H. Corrada Bravo**, B. Caffo, J.T. Leek<sup>\*</sup>, G. Parmigiani (2013). A decision-theory approach to interpretable set analysis for high-dimensional data. *Biometrics* 69 (3):614-23, doi:10.1111/biom.12060.
12. M.L. Nickerson<sup>^</sup>, K.M. Im, K.J. Misner, W. Tan, H. Lou, B. Gold, D.W. Wells, **H. Corrada Bravo**, K.M. Fredrikson, T.T. Harkins, P. Milos, B. Zbar, W.M. Linehan, M. Yeager, T. Andersson, M. Dean<sup>\*</sup>, G.S. Bova<sup>\*</sup> (2013). Somatic alterations contributing to metastasis of a castration-resistant prostate cancer. *Human Mutation* 34 (9): 1231-41, doi:10.1002/humu.22346.
13. **H. Corrada Bravo**<sup>^</sup>, V. Pihur, M. McCall, R.A. Irizarry, J.T. Leek<sup>\*</sup> (2012). Gene expression anti-profiles as a basis for cancer diagnostics. *BMC Bioinformatics* 13:272, doi:10.1186/1471-2105-13-272.
14. W. Shi<sup>^</sup>, G. Wahba, R.A. Irizarry, **H. Corrada Bravo**, S.J. Wright<sup>\*</sup> (2012). The Partitioned LASSO-Patternsearch Algorithm with Application to Gene Expression Data. *BMC Bioinformatics* 13:98, doi:10.1186/1471-2105-13-98.
15. T.S. Niranjana<sup>^</sup>, A. Adamczyk<sup>^</sup>, **H. Corrada Bravo**<sup>^</sup>, M. Taub, S.J. Wheelan, R.A. Irizarry, T. Wang<sup>\*</sup> (2011). Effective detection of rare variants in pooled DNA samples using Srfim and cross-pool tail-curve analysis. *Genome Biology* 12 (9):R93.
16. K. Hansen<sup>^</sup>, W. Timp<sup>^</sup>, **H. Corrada Bravo**<sup>^</sup>, S. Sabuncian<sup>^</sup>, B. Langmead<sup>^</sup>, O.G. McDonald, B. Wen, H. Wu, D. Diep, E. Briem, K. Zhang, R.A. Irizarry<sup>\*</sup>, A.P. Feinberg<sup>\*</sup> (2011). Increased methylation variation in epigenetic domains across cancer types. *Nature Genetics* 43 (8):768-75.
17. L. Shan<sup>^</sup>, H.C. Yang, S.A. Rabi, **H. Corrada Bravo**, J.D. Siliciano, R.A. Irizarry, H. Zhang, J. Margolick, R.F. Siliciano<sup>\*</sup> (2011). Influence of host gene transcription level and orientation on HIV-1 latency in a primary cell model. *Journal of Virology* 85 (11):5384-93.
18. A. Rivas<sup>^</sup>, K. Bohane, **H. Corrada Bravo**<sup>^</sup>, M. Tan, R. Tamargo, H.W. Francis (2011). A model for early prediction of facial nerve recovery after vestibular schwannoma surgery. *Otology & Neurotology* 32 (5):826-33.
19. H. Wu, R.A. Irizarry<sup>\*</sup>, **H. Corrada Bravo**<sup>^</sup> (2010). Intensity normalization improves color calling in SOLiD sequencing. *Nature Methods* 7:336-337.
20. J.T. Leek<sup>^</sup>, R. Scharpf, **H. Corrada Bravo**, D. Simcha, B. Langmead, W.E. Johnson, D. Geman, K. Baggerly, R.A. Irizarry<sup>\*</sup> (2010). Tackling the widespread and critical impact of batch effects in high-throughput data. *Nature Reviews Genetics* 11 (10):733-739.
21. **H. Corrada Bravo**<sup>^</sup>, R.A. Irizarry<sup>\*</sup> (2009). Model-based quality assessment and base-calling for second-generation sequencing data. *Biometrics* 66(3):665-74, doi:10.1111/j.1541-0420.2009.01353.x.
22. K. H. Eng<sup>^</sup>, **H. Corrada Bravo**<sup>^</sup>, S. Keles<sup>\*</sup> (2009). A phylogenetic mixture model for the evolution of gene expression. *Molecular Biology and Evolution* 26 (10):2363-2372.

23. **H. Corrada Bravo**<sup>^</sup>, K.E. Lee, B.E.K. Klein, R. Klein, S.K. Iyengar, G. Wahba\* (2009). Examining the relative influence of familial, genetic and environmental covariate information in flexible risk models. *Proceedings of the National Academy of Science* 106 (20): 8128-8133.
24. M. Acevedo<sup>^</sup>, T.M. Aide\*, L. J. Villanueva-Rivera, **H. Corrada Bravo**<sup>^</sup>, C. J. Corrada-Bravo\* (2009). Automated classification of bird and amphibian calls using machine learning: a comparison of methods. *Ecological Informatics* 4 (4):206-214.
25. A. T. Evan, R. Bennartz, V. Bennington, **H. Corrada Bravo**, A. K. Heidinger, N. M. Mahowald, C. S. Velde, G. Myhre, J. P. Kossin (2008). Ocean temperature forcing by aerosols across the Atlantic tropical cyclone development region. *Geochem, Geophys. Geosyst.* 9:Q05V04, doi:10.1029/2007GC001774.
26. C. Kuang, P. McMurry, A. McCormick, F. Eisele, S.H. Lee, L.H. Young, D.R. Benson, et al. (2007). A system for operational aerosol optical depth data assimilation over global oceans. *J. Geophys. Res* 113(D10):D10208.

## II.D Published Conference Proceedings

### II.D.1 Refereed Conference Proceedings

1. **H. Corrada Bravo**<sup>^</sup>, K. Eng, S. Keles, G. Wahba, S. Wright (2009). Estimating tree-structured covariance matrices via mixed integer programming. *Twelfth International Conference on Artificial Intelligence and Statistics (AISTATS '09); Journal of Machine Learning Research Workshop and Conference Proceedings*, 533:40.
2. **H. Corrada Bravo**<sup>^</sup>, R. Ramakrishnan (2007). Optimizing MPF queries: decision support and probabilistic inference. *26th ACM SIGMOD Intl. Conf. on Management of Data* 701:712.
3. **H. Corrada Bravo**<sup>^</sup>, D. Page, R. Ramakrishnan, J. Shavlik, V. Santos Costa (2005). A framework for set-oriented computation in inductive logic programming and its application in generalizing inverse entailment. *15th ILP Conf.* 69:86.

## II.E Conferences, Workshops and Talks

### II.E.2 Invited Talks

1. Exploring tumor epigenetic heterogeneity by cell-specific methylation pattern reconstruction. *CMU-Pitt Ph.D. Program in Computational Biology Seminar Series*, Pittsburgh, PA. April 2014.
2. Cell-specific methylation pattern reconstruction using minimum cost network flow algorithms. *Department of Mathematics, George Mason University*, Fairfax, VA. February 2014.
3. Gene expression anti-profiles as a basis for accurate universal cancer signatures. *Greenbaum Cancer Center, University of Maryland School of Medicine*, Baltimore, MD. September 2013.
4. Gene expression anti-profiles as a basis for accurate universal cancer signatures. *Institute for Genome Sciences, University of Maryland School of Medicine*, Baltimore, MD. March 2013.
5. Gene expression anti-profiles as a basis for accurate universal cancer signatures. *Department of Bioinformatics and Computational Biology, Genentech, Inc.*, South San Francisco, CA. February 2013.
6. Gene expression anti-profiles as a basis for accurate universal cancer signatures. *Innovation Center for Biomedical Informatics, Georgetown University*, Washington, DB. December 2012.

7. Statistical and computational methods for the analysis of pooled, targeted, second-generation re-sequencing data. *Biostatistics Department, University of Alabama*, Birmingham, AL. April, 2012.
8. Modeling gene expression variability for prediction in disease populations. *Department of Biostatistics, Columbia University School of Public Health*, New York, NY. December 2010.
9. Modeling gene expression variability for prediction in disease populations. *Johns Hopkins University School of Medicine*, Baltimore, MD. December 2010.
10. Gene expression variability in disease populations. *National Cancer Institute*, Bethesda, MD. October 2010.
11. Modeling uncertainty in second-generation sequencing data. *Dept. of Biostatistics, Harvard School of Public Health*, Boston, MA. November 2009.
12. Modeling and managing uncertainty in second-generation sequencing data. *Dept. of Computer Science, University of Maryland*, College Park, MD. October 2009.
13. Model-based quality assessment and base-calling for second-generation sequencing data. *University of Wisconsin-Milwaukee*, Milwaukee, MD. October 2009.
14. Model-based quality assessment and base-calling for second-generation sequencing data. *University of Manchester*, Manchester, England. October 2009.
15. Kernel methods for examining the relative influence of familial, genetic and environmental covariate information in risk models: results and (more importantly) extensions. *University of Wisconsin*, Madison, WI. May 2009.
16. Model-based quality assessment and base-calling for second-generation sequencing data. *University of Wisconsin*, Madison, WI. May 2009.
17. Model-based quality assessment and base-calling for second-generation sequencing data. *Case Western Reserve University*, Cleveland, OH. April 2009.
18. Data analysis at the computational/statistical sciences borderland: two examples from genomics. *Johns Hopkins Bloomberg School of Public Health*, Baltimore, MD. February 2009.
19. Estimating tree-structured covariance matrices via mixed-integer programming. *Johns Hopkins School of Public Health*, Baltimore, MD. January 2008.

### **II.E.3 Refereed Presentations**

1. Interactive and exploratory visualization of epigenome-wide data. *Joint Statistics Meetings*, Boston, MA. July 2014.
2. Interactive, Exploratory Visualization and Statistical Analysis of Genome-Scale Data. *International Biomteric Society ENAR Meeting*, Baltimore, MD. March 2014.
3. Gene expression anti-profiles as a basis for accurate universal cancer signatures. *ISMB '13*, Berlin, Germany. July 2013.
4. Srfim2: using basecalling model parameter estimates to understand sequencing bias. *2012 Joint Statistical Meetings*, San Diego, CA. August 2012.
5. Increased methylation variation in epigenetic domains across cancer types. *16th Annual International Conference on Research in Computational Molecular Biology (RECOMB)*, Barcelona, Spain. April 2012.

6. Statistical and computational methods for the analysis of pooled, targeted, second-generation re-sequencing data. *2011 Joint Statistical Meetings*, Miami Beach, FL.. August 2011.
7. Model-based quality assessment and base-calling for second-generation sequencing data. *WNAR/IMS annual meeting*, Portland, OR. June 2009.
8. Tuning regularized kernel estimation parameters for prediction. *SIAM Conference on Optimization*, Boston, MA. May 2008.
9. Optimizing MPF queries: decision support and probabilistic inference. *26th ACM SIGMOD Intl. Conf. on Management of Data*, Beijing, China. June 2007.
10. A framework for set-oriented computation in inductive logic programming and its application in generalizing inverse entailment.. *15th ILP Conf.*, Bonn, Germany. August 2005.

#### **II.E.7 Non-Refereed Presentations**

1. Epiviz(r): turning a genome browser into a display device. *Bioconductor conference*, Boston, MA. July 2014.
2. Statistical and computational methods for the analysis of pooled, targeted, second-generation re-sequencing data. *8th International Chinese Statistical Association (ICSA) International Conference*, Guangzhou, China. December 2010.
3. Model-based quality assessment and base-calling for second-generation sequencing data. *Conference on Next-Generation Sequencing*, Barcelona, Spain. October 2009.

#### **II.E.11 Symposia**

1. Gene expression network anti-profiles. *UMD/NIST Network Science Mini-Symposium*, College Park, MD. January, 2014.
2. Increased methylation variation in epigenetic domains across cancer types. *Omics Day, University of Maryland*, Shady Grove, MD. May 2012.
3. Model-based quality assessment and base-calling for second-generation sequencing data. *Third Annual Young Investigators Symposium on Genomics and Bioinformatics*, Baltimore, MD. September 2009.

#### **II.E.12 Workshops**

1. Gene expression network anti-profiles. *UMD/NIST Network Science Mini-Symposium*, College Park, MD. January, 2014.
2. Increased methylation variation in epigenetic domains across cancer types. *Omics Day, University of Maryland*, Shady Grove, MD. May 2012.
3. Model-based quality assessment and base-calling for second-generation sequencing data. *Third Annual Young Investigators Symposium on Genomics and Bioinformatics*, Baltimore, MD. September 2009.