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
- Affiliate Assistant Professor, Applied Math, Statistics and Scientific Computation
  University of Maryland, College Park, MD
  July 2011-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

- Resarch 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 Activites

II.C Articles in Refereed Journals

1. E. Alemu^, J.W. Carl, **H. Corrada Bravo***, S. Hannenhalli* (2014). Determinants of expression variability. *Nucleic Acids Research* 42 (6), 3503-14.

2. H.S. Parker^, **H. Corrada Bravo**, J.T. Leek* (2014). Removing batch effects for prediction problems with frozen surrogate variable analysis. *PeerJ* 2:e561, doi:10.7717/peerj.561.

3. W. Timp^, **H. Corrada Bravo**^, O.G. McDonald, M. Goggins, C. Umbricht, M. Zeiger, A.P. Feinberg*, R.A. Irizarry* (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.

4. F. Chelaru^#, L. Smith#, N. Goldstein#, **H. Corrada Bravo*** (2014). Epiviz: interactive visual analytics for epigenomics data. *Nature Methods* , doi:10.1038/nmeth.3038.

5. Pop, M.^, Walker, A.W., Paulson, J.#, 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.*, Nataro, J.P.* (2014). Diarrhea in young children from low-income countries leads to large-scale alterations in intestinal microbiota composition. *Genome Biology* 15, R76.

6. M. Aryee^, A. Jaffe, **H. Corrada Bravo**, C. Ladd-Acosta, A. Feinberg, K. Hansen*, R.A. Irizarry* (2014). Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA Methylation microarrays. *Bioinformatics* 30 (10):1363-9.

7. C. Ye^#, C. Hsiao#, **H. Corrada Bravo*** (2014). BlindCall: ultra-fast base-calling of second-generation sequencing by blind deconvolution. *Bioinformatics* 30 (9):1214-9.

8. N. Akula^, 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* (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. X. He^, R. Chatterjee^, S. John, **H. Corrada Bravo**, B.K. Sathyanarayana, S.C. Biddle, P.C. Fitzgerald, J.A. Stamatoyannopoulos, G.L. Hager, C. Vinson* (2013). Contribution of nucleosome binding preferences and co-occurring DNA sequences to transcription factor binding. *BMC Genomics* 14 (428).

10. M.L. Nickerson^, 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. Andresson, M. Dean*, G.S. Bova* (2013). Somatic alterations contributing to metastasis of a castration-resistant prostate cancer. *Human Mutation* 34 (9): 1231-41, doi:10.1002/humu.22346.

11. S. Boca^, **H. Corrada Bravo**, B. Caffo, J.T. Leek*, 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. J. Paulson[#], O.C. Stein, **H. Corrada Bravo***, M. Pop* (2013). Differential abundance analysis for microbial marker-gene surveys. *Nature Methods* 10 (12):1200-1202, doi:10.1038/nmeth.2658.

13. W. Shi^, G. Wahba, R.A. Irizarry, **H. Corrada Bravo**, S.J. Wright* (2012). The Partitioned LASSO-Patternsearch Algorithm with Application to Gene Expression Data. *BMC Bioinformatics* 13:98, doi:10.1186/1471-2105-13-98.

14. **H. Corrada Bravo**^, V. Pihur, M. McCall, R.A. Irizarry, J.T. Leek* (2012). Gene expression anti-profiles as a basis for cancer diagnostics. *BMC Bioinformatics* 13:272, doi:10.1186/1471-2105-13-272.

15. T.S. Niranjan^, A. Adamczyk^, **H. Corrada Bravo**^, M. Taub, S.J. Wheelan, R.A. Irizarry, T. Wang* (2011). Effective detection of rare variants in pooled DNA samples using Srfim and cross-pool tail-curve analysis. *Genome Biology* 12 (9):R93.

16. A. Rivas^, K. Bohane, **H. Corrada Bravo**^, 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.

17. L. Shan^, H.C. Yang, S.A. Rabi, **H. Corrada Bravo**, J.D. Siliciano, R.A. Irizarry, H. Zhang, J. Margolick, R.F. Siliciano* (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. K. Hansen^, W. Timp^, **H. Corrada Bravo**^, S. Sabunciyan^, B. Langmead^, O.G. McDonald, B. Wen, H. Wu, D. Diep, E. Briem, K. Zhang, R.A. Irizarry*, A.P. Feinberg* (2011). Increased methylation variation in epigenetic domains across cancer types. *Nature Genetics* 43 (8):768-75.

19. H. Wu, R.A. Irizarry*, **H. Corrada Bravo**^(2010). Intensity normalization improves color calling in SOLiD sequencing. *Nature Methods* 7:336-337.

20. J.T. Leek^, R. Scharpf, **H. Corrada Bravo**, D. Simcha, B. Langmead, W.E. Johnson, D. Geman, K. Baggerly, R.A. Irizarry* (2010). Tackling the widespread and critical impact of batch effects in high-throughput data. *Nature Reviews Genetics* 11 (10):733-739.

21. M. Taub^, **H. Corrada Bravo**, R.A. Irizarry* (2010). Overcoming bias and systematic errors in next generation sequencing data. *Genome Medicine* 2 (12):87.

22. M. Acevedo^, T.M. Aide*, L. J. Villanueva-Rivera, **H. Corrada Bravo**^, 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.

23. K. H. Eng^, **H. Corrada Bravo**^, S. Keles* (2009). A phylogenetic mixture model for the evolution of gene expression. *Molecular Biology and Evolution* 26 (10):2363-2372.

24. **H. Corrada Bravo^**, 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.

25. **H. Corrada Bravo^**, R.A. Irizarry* (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.

26. 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.

27. 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^**, 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, 5*33:40.

2. **H. Corrada Bravo^**, 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^**, 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 Franciso, 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.

II.F Professional Publications

II.F.2 Pre-print/Working Paper (Not Work in Progress)

1. W. Dinalankara^#, **H. Corrada Bravo** (2013). Anomaly classification with the anti-profile support vector machine. *arXiv preprint server* arXiv:1301.3514 [stat.ML].

2. H.S. Parker^, **H. Corrada Bravo**, J.T. Leek (2013). Removing batch effects for prediction problems with frozen surrogate variable analysis. *arXiv preprint server* arXiv:1031.3947 [stat.ME].

3. S. Boca^, **H. Corrada Bravo**, B. Caffo, J.T. Leek, G. Parmigiani (2010). A decision-theory approach to interpretable set analysis for high-dimensional data. *Johns Hopkins University, Dept. of Biostatistics Working Papers* Working paper 211. http://biostats.bepress.com/jhubiostat/paper211.

4. R.A. Irizarry, **H. Corrada Bravo**^(2009). Model-based quality assessment and base-calling for second-generation sequencing data. *Johns Hopkins University, Dept. of Biostatistics Working Papers* Working paper 184. http://biostats.bepress.com/jhubiostat/paper184.

II.H Completed Creative Works

II.H.8 Software and Applications

1. Epiviz: Interactive visualization for genomics data.
   Released June 2013. http://epiviz.cbcb.umd.edu.

2. Epivizr: interactive visualization of genomics data in R/Bioconductor.
   Released June 2013. http://bioconductor.org/packages/release/bioc/html/epivizr.html.

3. antiProfiles: Gene expression anti-profiles as a basis for accurate universal cancer signatures.
   Released April 2013. http://bioconductor.org/packages/release/bioc/html/antiProfiles.html.

4. Healthvis: Interactive visualization in health.
   Released April 2013. http://healthvis.org.

5. Minfi: Analyze Illumina's 450k methylation arrays.
   Released Febraury 2013. http://bioconductor.org/packages/release/bioc/html/minfi.html.

6. metagenomeSeq: Statistical analysis for sparse high-throughput sequencing.
   Released February 2013. http://bioconductor.org/packages/release/bioc/html/metagenomeSeq.html.

7. bumphunter: Tools for finding bumps in genomic data.
   Released February 2013. http://bioconductor.org/packages/release/bioc/html/bumphunter.html.

8. Servic4e: Effective detection of rare variants in pooled DNA samples.
   Released November 2011. http://www.cbcb.umd.edu/~hcorrada/secgen.

9. Rsolid: Intensity normalization for SOLiD sequencing.
   Released April 2010. http://www.cbcb.umd.edu/~hcorrada/secgen.

10. Srfim: Model-based base-calling and quality assessment for second-generation sequencing.
    Released April 2010. http://www.cbcb.umd.edu/~hcorrada/secgen.

11. Rcsdp: An interface to the CSDP semidefinite programming library for R.
    Released Dec. 2008. http://cran.r-project.org/web/packages/Rcsdp/index.html.

12. Rcplex: An interface to the CPLEX optimization engine for R.
Released Jan. 2008. http://cran.r-project.org/web/packages/Rcplex/index.html.

II.H.9 Websites

1. Epiviz: Interactive visualization for genomics data.
Released May 2014. http://epiviz.cbcb.umd.edu/help.

II.J Sponsored Research

II.J.1 Grants

1. Title: R01: Analysis tools and software for second generation sequencing
Funding Agency: NIH
Recipient Institution: Johns Hopkins University, Rafael A. Irizarry (PI)
Amount Awarded: $380,400 subcontracted to UMCP ($1,230,000 total awarded by NIH)
Dates: August 2010-May 2013
Role: PI

2. Title: R01: Alignment software for second generation sequencing
Funding Agency: NIH
Recipient Institution: Johns Hopkins University, Steven Salzberg (PI)
Amount Awarded: $432,259.00 subcontracted to UMCP ($3,585,436 total awarded by NIH)
Dates: May 2011-April 2014
Role: PI

II.K Fellowships, Gifts and Other Funded Research

II.K.1 Fellowships

1. Ford Fellowship, National Academies of Science.

2. Advanced Opportunity Fellowship, University of Wisconsin-Madison.

II.L Submissions and Works in Progress

II.L.1 Current Grant Applications

1. Title: CAREER: Data Science in Genomics: visualizing and learning from inferred features
Funding Agency: NSF
Recipient Institution: University of Maryland, College Park
Amount Requested: $634,198
Dates: June 2015-May 2020
Role: PI

2. Title: R01: Analysis tools and software for second generation sequencing
Funding Agency: NIH
Recipient Institution: Dana Farber Cancer Institute, Rafael A. Irizarry (PI)
Amount Requested: $397,903 subcontracted to UMCP
Dates: September 2014-August 2018
Role: PI

3. Title: R01: Integrative visual and computational exploratory analysis of genomics data
   Funding Agency: NIH
   Recipient Institution: University of Maryland, College Park
   Amount Requested: $1,222,398
   Dates: April 2015-March 2018
   Role: PI

4. Title: NRT-DESE; Network biology: from data to information to insights
   Funding Agency: NSF
   Recipient Institution: University of Maryland, College Park
   Amount Requested: $2,997,748
   Dates: September 2015-August 2020
   Role: co-PI

II.L.2 Manuscripts in Preparation

II.L.3 Manuscripts under Review

1. H. Talukder^[#], J. Paulson[#], **H. Corrada Bravo***. Finding regions of interest in high throughput genomics data using smoothing splines. *BMC Bioinformatics, submitted.*

2. W. Huber^, V. Carey, R. Gentleman, M. Carlson, B.S. Carvalho, **H. Corrada Bravo**, S. Davis, L. Gatto, T. Girke, R. Gottardo, F. Hahne, K. Hansen, R.A. Irizarry, M. Lawrence, M.I. Love, J. MacDonald, V. Obenchain, A.K. Olés, H. Pagés, P. Shannon, G. Smyth, D. Tenenbaum, L. Waldron, M. Morgan*. Orchestrating high-throughput genomic analysis with Bioconductor. *Nature Methods, submitted.*

3. H. Talukder^[#], **H. Corrada Bravo**, Z. Dezman*, B. Golden, S. Mankad^. Does health insurance matter? Establishing insurance states as a risk factor for mortality rate. *Injury, submitted.*

4. W. Dinalankara^[#], **H. Corrada Bravo***. Reproducible tumor diagnosis and prognosis signatures via gene expression anti-profiles. *Bioinformatics, submitted.*

5. K. Okrah^[#], **H. Corrada Bravo***. Shape analysis for high-throughput transcriptomics experiment data. *Biostatistics*, *revising to resubmit.

II.L.4 Working papers in Preparation

papers in preparation

II.L.5 Designs in preparation

software in prep