

Recent publications and research funding

I. Selected Research, Scholarly, Creative and Professional Activities during Past Five Years

I.A. Articles in Refereed Journals

1. L. Dillon[^], K. Okrah^{^#}, V. Hughitt, R. Suresh, Y. Li, M. Fernandes, A. Belew, **H. Corrada Bravo**, D. Mosser, N. El-Sayed* (2015). Transcriptomic profiling of gene expression and RNA processing during <i>Leishmania major</i> differentiation. *Nucleic Acids Research* 43 (14): 6799-6813, doi:10.1093/nar/gkv656.
2. W. Dinalankara^{^#}, **H. Corrada Bravo*** (2015). Gene expression signatures based on variability can robustly predict tumor progression and prognosis. *Cancer Informatics* 14:71-81.
3. 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* (2015). Orchestrating high-throughput genomic analysis with Bioconductor. *Nature Methods* 12 115-121, doi:10.1038/nmeth.3252.
4. K. Okrah^{^#}, **H. Corrada Bravo*** (2015). Shape analysis for high-throughput transcriptomics experiment data. *Biostatistics* , doi:10.1093/biostatistics/kxv018.
5. 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.
6. 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.
7. F. Chelaru^{^#}, L. Smith[#], N. Goldstein[#], **H. Corrada Bravo*** (2014). Epiviz: interactive visual analytics for epigenomics data. *Nature Methods* 11 938-940, doi:10.1038/nmeth.3038.
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. 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.
10. 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.
11. E. Alemu[^], J.W. Carl, **H. Corrada Bravo***, S. Hannenhalli* (2014). Determinants of expression variability. *Nucleic Acids Research* 42 (6), 3503-14.

12. 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., Staes, 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.
13. 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.
14. 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).
15. 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.
16. 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. Andersson, 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.
17. **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.
18. 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.
19. T.S. Niranjana[^], 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.
20. K. Hansen[^], W. Timp[^], **H. Corrada Bravo**[^], S. Sabuncuyan[^], 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.
21. 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.
22. 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.
23. 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.
24. M. Taub[^], **H. Corrada Bravo**, R.A. Irizarry* (2010). Overcoming bias and systematic errors in next generation sequencing data. *Genome Medicine* 2 (12):87.
25. H. Wu, R.A. Irizarry*, **H. Corrada Bravo**[^](2010). Intensity normalization improves color calling in SOLiD sequencing. *Nature Methods* 7:336-337.
26. 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.

27. **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.
28. **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.
29. 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.
30. 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.
31. 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.

I.B. Published Conference Proceedings

I.B.1. Refereed Conference Proceedings

1. F. Chelaru^{^*#}, **H. Corrada Bravo**^{*} (2015). Epiviz: a view inside the design of an integrated visual analysis software for genomics. *BioVis 2015; BMC Bioinformatics* 16(Supl 11):S4.
2. **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*, 533:40.
3. **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.
4. **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.

I.C. Completed Creative Works

I.C.1. 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 February 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>.

I.C.2. Websites

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 February 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>.

I.D. Sponsored Research

I.D.1. Grants

1. Title: R01: Analysis tools and software for second generation sequencing
Funding Agency: NIH HG005220
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 HG006102
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
3. Title: R01: Overcoming bias and unwanted variability in next generation sequencing
Funding Agency: NIH HG005220
Recipient Institution: Dana Farber Cancer Institute, Rafael A. Irizarry (PI)
Amount Awarded: \$397,903 subcontracted to UMCP
Dates: March 2015-February 2019
Role: PI
4. Title: R01: Integrative visual and computational exploratory analysis of genomics data
Funding Agency: NIH
Recipient Institution: University of Maryland, College Park
Amount Awarded: \$1,800,421
Dates: September 2015-August 2019
Role: PI

I.E. Submissions and Works in Progress

I.E.1. Current Grant Applications

1. Title:
Funding Agency:
Recipient Institution:
Amount Requested:
Dates:
Role:

I.E.2. Manuscripts under Review

1. S. Braid[^], K. Okrah[#], A. Shetty, **H. Corrada Bravo**^{*}. DNA methylation differences in cord blood between term and preterm infants is associated with differences in cell composition. *Clinical Epigenetics*, revision submitted.

2. M. Sharmin[^], **H. Corrada Bravo**^{*}, S. Hannenhalli^{*}. Characterizing the regulatory landscape of large hypomethylated blocks in colon cancer. *BMC Cancer*, *revision submitted*.
3. L.A. Dillon[^], R. Suresh, K. Okrah[#], **H. Corrada Bravo**, D.M. Mosser, N.M. El-Sayed^{*}. Simultaneous transcriptional profiling of *Leishmania major* and its murine macrophage host cell reveals insights into host-pathogen interactions. *BMC Genomics*, *revision submitted*.
4. Y. Li[^], K.L. Caradonna[^], K. Okrah^{^#}, J. Choi, P. Padmanabhan, D.M. Ndegwa, S. Shah-Simpson, M.R. Temanni, **H. Corrada Bravo**, N. El-Sayed^{*}, B.A. Burleigh^{*}. Transcriptome Remodeling in *Trypanosoma cruzi* and Human Cells during Intracellular Infection. *PLOS Pathogens*, *revising to resubmit*.
5. J. Wagner^{^#}, J. Paulson[#], X.S. Wang, B. Bhattacharjee, **H. Corrada Bravo**^{*}. Privacy preserving human microbiome analysis using secure computation. *Bioinformatics*, *revising to resubmit*.
6. M. Sharmin^{^#}, **H. Corrada bravo**, S. Hannenhalli^{*}. Heterogeneity of TF binding specificity models within and across cell lines. *Genome Research*, *submitted*.
7. M.C. Fernandes[^], L.A. Dillon[^], A.T. Belew, **H. Corrada Bravo**, D.M. Mosser, N. El-Sayed^{*}. Dual transcriptome profiling of *Leishmania* parasites and the human macrophage. *Cell Host & Microbe*, *submitted*.
8. F. Dorri^{^#}, L. Mendelowitz, **H. Corrada Bravo**^{*}. methylFlow: cell-specific methylation pattern reconstruction from high-throughput bisulfite-converted DNA sequencing. *HitSEQ 2015; Bioinformatics*, *revision under review*.