|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| BIOGRAPHICAL SKETCH Provide the following information for the Senior/key personnel and other significant contributors. Follow this format for each person.  **DO NOT EXCEED FOUR PAGES.** | | | | |
|  | | | | |
| NAME  Héctor Corrada Bravo | | POSITION TITLE  Assistant Professor of Computer Science | | |
| eRA COMMONS USER NAME (credential, e.g., agency login)  HCORRADA | |
| EDUCATION/TRAINING *(Begin with baccalaureate or other initial professional education, such as nursing, include postdoctoral training and residency training if applicable.)* | | | | |
| INSTITUTION AND LOCATION | DEGREE  *(if applicable)* | | MM/YY | FIELD OF STUDY |
| Johns Hopkins University | B.M. | | 05/97 | Music |
| University of Wisconsin, Madison | Ph.D. | | 08/08 | Computer Science |
| Johns Hopkins University | Postdoctoral | | 06/10 | Biostatistics-Computational Genomics |
|  |  | |  |  |
|  |  | |  |  |

A. Personal Statement

I have received interdisciplinary training in Computer Science, Biostatistics and Computational Genomics. I have over 7 years of experience and a broad background in computational genomics and biostatistics, with specific experience in the analysis of high-throughput epigenetic and metagenomics data. I lead development of the Epiviz (<http://epiviz.cbcb.umd.edu>, Chelaru et al., 2014) big genomics data visualization tool. I also lead development in new statistical and computational methods for the analysis of metagenomics data, and its associated metagenomeSeq software tool (Paulson, et al., 2013). I collaborate extensively with biomedical researchers. I have developed methods and software for second-generation sequencing data analysis: quality assessment, base-calling, variant discovery, bisulfite sequencing for DNA methylation measurement, and RNAseq analysis as co-Investigator on several NIH-funded grants. I have overseen training and advised pre-doctoral students in Computer Science, Applied Statistics, Scientific Computation and Computational Biology, who have transitioned to post-doctoral positions both in academia and industry. In summary, my demonstrated experience and expertise in developing and applying big data analysis tools and methods, and record of training of pre-doctoral students in this area qualifies me to participate in the proposed training activities of this project.

B. Positions and Honors

Positions and Employment

2003-2008 Research Assistant, Departments of Computer Sciences and Statistics, University of Wisconsin, Madison, WI

2008-2010 Postdoctoral Fellow, Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD

2010- Assistant Professor, Department of Computer Science, UM Institute for Advanced Computer Studies and the Center for Bioinformatics and Computational Biology, University of Maryland, College Park, MD

Other Experience and Professional Memberships

2010- Member, Association for Computing Machinery

2010- Member, International Society for Computational Biology

Honors

2003-2007 Ford Fellowship, National Academies of Science

2004-2005 Advanced Opportunity Fellowship, University of Wisconsin-Madison

C. Contribution to Science

1. *Cancer Functional Genomics.* Understanding the role of epigenetics in tumorigenesis is a fundamental question in cancer research. I was co-first author in a landmark publication in cancer epigenomics: the first whole methylome study at base-pair resolution in cancer using second-generation sequencing. This study was the first to identify long regions in the genome where stable methylation in the normal cell is lost in tumors, and identifying resulting changes in gene expression. Subsequent studies have identified similar mechanisms as part of one of the first studies of DNA methylation across multiple tissue types and across cancer progression. I have led subsequent studies on the impact of these findings on gene expression regulation and how to translate these to stable expression-based biomarkers.

1. Hansen, K., Timp, W., **Bravo, H.C.,** Sabunciyan, S., Langmead, B., McDonald, O.G., Wen, B., Wu, H., Diep, D., Briem, E., Zhang, K., Irizarry, R.A., Feinberg, A.P. (2011). Increased methylation variation in epigenetic domains across cancer types. Nature Genetics 43 (8), 768-75. PMC3145050.
2. Timp, W., **Bravo, H. C**., McDonald, O. G., Goggins, M., Umbricht, C., Zeiger, M., et al. (2014). Large hypomethylated blocks as a universal defining epigenetic alteration in human solid tumors. Genome Medicine, *6*(8), 61. doi:10.1186/s13073-014-0061-y
3. **Bravo, H.C.**, 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. PMC3487959.
4. Dinalankara, W., **Bravo, H.C.** (2015). Gene expression signatures based on variability can robustly predict tumor progression and prognosis. Cancer Informatics, *in press.*
5. **Bravo, H.C.,** Irizarry, R.A., Leek, J.T. (2012). antiProfiles: Implementation of gene expression anti-profiles*.* Software, <http://bioconductor.org/packages/release/bioc/html/antiProfiles.html>. doi:10.5281/zenodo.17260

2. *Statistical and Computational Methods for High-Throughput Genomics.* Proper processing and modeling of high-dimensional data like sequencing or array-based assays of gene expression or DNA methylation is essential to proper understanding of the mechanisms measured. My group has developed novel, robust methods for the preprocessing and modeling of DNA methylation using the Illumina HumanMethylation450k array, commonly used as part of Epigenome-wide association studies. My group has also developed new methodology for the modeling of gene expression variability allowing understanding of how this phenotype relates to genomic signals in normal tissue. All of these methods are implemented in software packages that are open source, distributed and used widely. My group has also developed interactive visualization tools that make the use of these statistical methods more effective in exploratory and collaborative analysis, while making dissemination obtained via these methods richer and reproducible.

1. Chelaru, F., Smith, L., Goldstein, N., **Bravo, H.C.**, (2014). Epiviz: interactive visual analytics for functional genomics data. Nature Methods, 11 (9)*.* PMC Journal, in process.
2. M. Aryee, A. Jaffe, **H.C. Bravo,** C. Ladd-Acosta, A. Feinberg, K. Hansen, R.A. Irizarry (2014). Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infimium DNA methylation microarrays. *Bioinformatics,* 30 (10), 1363-9. PMC4016708.
3. Leek, J.T., Scharpf, R., **Bravo, H.C.**,Simcha, D., Langmead, B., Johnson, E., Geman, D., Baggerly, K., & Irizarry, R.A. (2010). Tackling the widespread and critical impact of batch effects in high-throughput data. Nature Reviews Genetics*, 11 (10),* 733-739. PMID: 20838408.
4. Alemu, E., Carl, J.W., **Bravo**, **H.C.,** Hannenhalli, S. (2014). Determinants of expression variability. *Nucleic Acids Research,* 42 (6), 3503-3514. PMC3973347.
5. Chelaru, F. Smith, L., Goldstein, N., **Bravo, H.C.** (2014). Epiviz: interactive visual analytics for functional genomics data. Software, <https://github.com/epiviz/epiviz/>. doi: 10.5281/zenodo.17261.

3. *Statistical and Computational Methods for High-Throughput Metagenomics.* Large epidemiological studies using sequencing methods to probe microbial communities are becoming commonplace powerful methods of understanding the role of microbe-host interactions in health and disease. My group has developed statistical methods designed specifically to pre-process and model these data to effectively and robustly detect taxonomic units associated with phenotypes of interest. These methods were designed to analyze data from one the earliest large epidemiological studies using metagenomic data. These methods are implemented and distributed as open source software, used frequently by other groups as part of standard metagenomic analysis pipelines.

1. Paulson, J., Stein, O.C., **Bravo, H.C.,** Pop, M., (2014). Differential abundance analysis for microbial marker-gene surverys. Nature Methods, 10 (12): 1200-1202. PMC4010126.
2. Pop, M., Walker, A.W., Paulson, J., Lindsey, B., Antonio, M., Hossain, M.A., Oundo, J., Tamboura, B., Mai, V., Astrovskaya, I., **Bravo, H.C.,** 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(6), p. R76*.* PMC Journal, in process.
3. Paulson, J., Pop, M., **Bravo, H.C.** (2014). metagenomeSeq: statistical analysis for sparse high-throughput sequencing. Software, <http://bioconductor.org/packages/release/bioc/html/metagenomeSeq.html>.

4. *Statistical and Computational Methods for preprocessing and quality assessment of high-throughput sequencing data.* Technological advances in in high-throughput sequencing have revolutionized biomedical research. The use of these technologies for multiple downstream analyses, including variant analysis is predicated on the robust processing of raw signal to create the DNA sequence reads used for downstream analysis. Our group has developed statistical methods and software to efficiently and robustly perform this crucial upstream analysis step. We have also studied the effect of biases in these technologies in downstream analyses critical to the clinical translation of findings made with these technologies.

1. **Bravo, H.C.** & Irizarry, R.A. (2009). Model-based quality assessment and base-calling for second-generation sequencing data. Biometrics 66(3), 665-74. PMC3439029.
2. M. Taub, **H. Corrada Bravo**, R.A. Irizarry (2010). Overcoming bias and systematic errors in next generation sequencing data. Genome Medicine2(12):87. PMC3025429.
3. Wu, H., Irizarry, R.A. & **Bravo, H.C.** (2010).Intensity normalization improves color calling in SOLiD sequencing. Nature Methods 7, 336-337. PMC3142576.
4. Niranjan, T.S., Adamaczyk, A., **Bravo, H.C.**, Taub, M., Wheelan, S.J., Irizarry, R.A., Wang, T (2011). Effective detection of rare variants in pooled DNA samples using Srfim and cross-pool tail-curve analysis. Genome Biology 12 (9), R93. PMC3439029.
5. Ye, C., Hsiao, C., **Bravo, H.C.** (2014). BlindCall: ultra-fast base-calling of second-generation sequencing by blind deconvolution. Bioinformatics 30 (9), 1214-1219. PMC3998134.

5. *Graph-based methods for structured data analysis and integration.* My dissertation work was focused in developing statistical and computational methods for the analysis of structured data. First, we worked on data integration using kernel methods. These are flexible statistical methods that permit integration of both standard observation-specific data and relationships between observations. We used these methods to analyze the relative influence of familial, genetic and environmental covariate information in disease risk models. The second aspect was developing structured covariance models used to understand evolution of gene expression in closely related species. We also used a related method to understand the genetic relationship between metastases and the primary tumor in pancreatic cancer.

1. **Bravo, H.C.,** Lee, K.E., Klein, B.E.K., Klein, R., Iyengar, S.K. & Wahba, G. (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. PMC2677979.
2. **Bravo, H.C.,** Eng, K.H., Keles, S. Wahba, G., & Wright, S. (2009). Estimating tree-structured covariance matrices via mixed-integer programming. Journal of Machine Learning Research Workshop and Conference Proceedings (AISTATS ’09), 5, 33-40. PMC3212858.
3. Eng, K.H., **Bravo, H.C.,** & Keles, S. (2009). A phylogenetic mixture model for the evolution of gene expression. Molecular Biology and Evolution, 26(10), 2363-2372. PMC2738779.
4. M.L. Nickerson, K.M. Im, K.J. Misner, A.L. Yates, D.W. Wells, **H. Corrada Bravo**, K. Fredrikson, W. Tan, D. Research Support

**Ongoing Research Support**

**Completed Research Support**

R01 HG005220-03 R.A., Irizarry (PI) 08/01/2010-05/25/2014

Analysis tools and software for second generation sequencing

The goal of this project is to develop methods and software for analysis of second-generation data. This includes methods for quality assessment, and for analysis of whole-epigenome data.

Role: Co-Investigator

R01 HG006102-02 S. Salzberg (PI) 05/01/2011-04/25/2014

Alignment software for second-generation sequencing

The goal of this project is to develop software for alignment of short read data produced by second generation sequencing methods.

Role: Co-Investigator