Samantha Zarate

slzarate@jhu.edu • GitHub: slzarate • in: slzarate

Education

Current *Aug 2019*	Ph.D. Student in Computational Biology, **Schatz Lab** Department of Computer Science, **Johns Hopkins University** Relevant classes: Computational Genomics: Sequences, Sketching and Indexing for Sequences, Applied Comparative Genomics, Machine Learning: Deep Learning
June 2016 *Sept 2012*	Bachelor of Science in Biomedical Computation, **Stanford University** Completed a Notation in Science Communication, creating an ePortfolio of artifacts depicting growth and development in the realm of science communication.

Work and Research Experience

Aug 2019 *Aug 2016*	Scientist at **DNAnexus**, Mountain View, CA • Implemented popular bioinformatics tools on the DNAnexus platform, including structural variation tools such as Parliament2 • Documentation co-lead, including structuring the content, writing and editing pages, and helping prioritize product documentation; co-created a new site for DNAnexus documentation • Provided scientific and bioinformatics services for all customers in the country of Singapore, including the Genome Institute of Singapore and the National University of Singapore • Worked with various other customers to support, accelerate, and productionize their bioinformatics pipelines • Mentored a summer intern, helping her develop and productionize a DNAnexus-recommended RNA-seq analysis workflow
Sept 2015 *June 2015*	Bioinformatics Intern at **DNAnexus**, Mountain View, CA • Productionized and parallelized tools for the DNAnexus platform including WGSA, CNVnator, and SomaticSniper • Productionized and parallelized structural variant calling and analysis tools for the DNAnexus platform such as Parliament, Lumpy, Delly, and Pindel • Attended the PacBio SMRT Informatics developers' conference as well as the Genome in a Bottle conference, both at the National Institute for Standards and Technology in Gaithersburg, MD, where work on Parliament was presented
June 2015 *Apr 2013*	Research Assistant at **Luo Lab**, Stanford University, Stanford, CA Understanding the function of *Rai1* • Using Mosaic Analysis with Double Markers (MADM) in mice, conducted dendritic spine density analysis and cell quantification in various regions of the brain • Analyzed *Rai1* expression in embryonic carcinoma murine stem cells differentiated into neurites and muscle cells and differential expression of *Rai1* using RNA-seq transcription and microarray data • Conducted bioinformatics analysis of RNA-Seq and microarray data

Preprints & Manuscripts

M1. A diploid assembly-based benchmark for variants in the major histocompatibility complex. C-S Chin, J. Wagner, Q. Zeng, E. Garrison, S. Garg, A. Fungtammasan, M. Rautiainen, S. Aganezov, M. Kirsche, **S. Zarate**, M. C. Schatz, C. Xiao, W. J. Rowell, C. Markello, J. Farek, F. J. Sedlazeck, V. Bansal, B. Yoo, N. Miller, X. Zhou, A. Carroll, A. Martinez Barrio, M. Salit, T. Marschall, A. T. Dilthey, J. M. Zook (2020) *Nature Communications* doi: `https://doi.org/10.1038/s41467-020-18564-9`

P3. Benchmarking challenging small variants with linked and long reads. J. Wagner, N. D Olson, L. Harris, Z. Khan, J. Farek, M. Mahmoud, A. Stankovic, V. Kovacevic, A. M. Wenger, W. J. Rowell, C. Xiao,

B. Yoo, N. Miller, J. A. Rosenfeld, B. Ni, **S. Zarate**, M. Kirsche, S. Aganezov, M. Schatz, G. Narzisi, M. Byrska-Bishop, W. Clarke, U. S. Evani, C. Markello, K. Shafin, X. Zhou, A. Sidow, V. Bansal, A. Martinez Barrio, I. T. Fiddes, A. Fungtammasan, C-S Chin, F. J. Sedlazeck, A. Carroll, M. Salit, J. M. Zook (2020) *bioRxiv* doi: `https://doi.org/10.1101/2020.07.24.212712`

P2. Multiethnic catalog of structural variants and their translational impact for disease phenotypes across 19,652 genomes. F. J. Sedlazeck, B. Yu, A. J. Mansfield, H. Chen, O. Krasheninina, A. Tin, Q. Qi, **S. Zarate**, J. L. Traynelis, V. Menon, BCM HGSC Sequencing Lab, J. Hu, H. Doddapaneni, G. A. Metcalf, J. Coresh, R. C. Kaplan, D. M. Muzny, G. Jun, R. A. Gibbs, W. J. Salerno, E. Boerwinkle (2020) *bioRxiv* doi: `https://doi.org/10.1101/2020.05.02.074096`

P1. Parliament2: Fast Structural Variant Calling Using Optimized Combinations of Callers. **S. Zarate**, A. Carroll, O. Krasheninina, F. J. Sedlazeck, G. Jun, W. Salerno, E. Boerwinkle, R. Gibbs (2018) *bioRxiv* doi: `https://doi.org/10.1101/424267`

Posters

- Parliament2: Accurate Structural Variant Calling at Scale. **S. Zarate**, A. Carroll, M. Mahmoud, O. Krasheninina, G. Jun, W. J. Salerno, M. C. Schatz, E. Boerwinkle, R. A. Gibbs, F. Sedlazeck (2020) Biological Data Science, Cold Spring Harbor, NY (virtual).
- Parliament2: Fast Structural Variant Calling Using Optimized Combinations of Callers. **S. Zarate**, A. Carroll, O. Krasheninina, F. Sedlazeck, G. Jun, W. J. Salerno, E. Boerwinkle, R. A. Gibbs (2019) Advances in Genome Biology and Technology (AGBT), Marco Island, FL.
- Parliament2: Benchmarking a Structural Variant Consensus Caller Compared to Individual Methods. A. Carroll, **S. Zarate**, W. J. Salerno, F. Sedlazeck, O. Krasheninina, R. Gibbs (2018) 23andMe Genome Research Day, Mountain View, CA.
- Parliament2: Benchmarking a Structural Variant Consensus Caller Compared to Individual Methods. A. Carroll, **S. Zarate**, W. J. Salerno, F. Sedlazeck, O. Krasheninina, R. Gibbs (2018) Biology of Genomes, Cold Spring Harbor, NY.
- CHORUS: Enabling the sharing of federated genomic data associated with subjects from varied Singapore ethnic groups. **S. Zarate**, J. Ow, Y. Hwang, Y. Men, P. Lim, P. Ng (2017) Biology of Genomes, Cold Spring Harbor, NY.
- Analyzing the function of Smith Magenis Syndrome-associated gene *Rai1*. **S. Zarate**, C. Guenthner, L. Luo (2014) SURPS (Symposium of Undergraduate Research and Public Service) poster session, Stanford, CA.
- Understanding the function of RAI1 through genetic mosaic analysis in mice. **S. Zarate**, C. Guenthner, L. Luo (2013) Bio-X Undergraduate Fellow poster session, Stanford, CA.

Skills

- **Scientific:** R, SciPy, NumPy, MATLAB, LATEX, popular bioinformatics software (e.g. BWA-MEM, Samtools, GATK, etc.), structural variant callers (e.g. Lumpy, Breakdancer, Delly, etc.)
- **Technical:** Python, Bash, Docker, WDL, C, C++, Java, version control (GitHub, Git, Mercurial), cloud software (AWS, Google Cloud), DNAnexus, AnVIL, PyTorch
- **Other:** Scientific communication and technical writing, moderate grasp of Spanish