Bari J. Ballew, Ph.D.

https://github.com/bballew

(973) 713-6549
bari.ballew@gmail.com

17405 Buehler Road
Olney, MD 20832

Experienced human genomics scientist with in-depth knowledge of DNA sequencing data analysis. Proven track record in process improvement and production-quality, reproducible bioinformatic pipeline development. Excellent cross-discipline communication skills.

PROFESSIONAL EXPERIENCE

Bioinformatics Scientist
Cancer Genomics Research Laboratory, NCI/Leidos Biomedical Research, Inc. 01/2017-present

- Led our microbiome pipeline development team; instituted best practices including version control, code review, and unit/regression/black box testing
- Wrote a modular, platform-agnostic pipeline to coordinate structural variant calling using multiple containerized callers, then annotate and compare results across callers; awarded $20,000 for development and deployment to cloud platforms
- Developed a pipeline to implement non-negative matrix factorization to detect mutational signatures in somatic sequencing data; discovered evidence for a novel risk factor in esophageal cancer via the signatures detected
- Designed an ensemble germline variant calling pipeline using GATK HaplotypeCaller and Google's DeepVariant, using GLnexus to merge gVCFs over large cohorts; combined existing VCF merging tools with custom Python scripts to generate an output VCF that maintains all tags from the original two callers
- Worked closely with Sentieon to evaluate their software's performance in our environment and establish usage guidelines and documentation
- Managed one direct report; introduced a productivity dashboard to track my team's workload and progress toward annual goals
- Taught workshops on pipeline development with Snakemake and on data analysis with Python's pandas package (see GitHub repositories for interactive tutorials)

Genome Scientist
Personal Genome Diagnostics, Baltimore, MD 01/2016-12/2016

- Developed, implemented, and maintained automated report generation tool, eliminating sources of manual error and reducing turnaround time for reporting NGS results from hours to seconds
- Wrote Perl scripts to facilitate extraction and formatting of required data from raw pipeline output to concise, human-readable summary reports
- Collaborated closely with software engineers to establish requirements, regression testing strategy, and user acceptance testing plans for CLIA reports and related software
- Evaluated the effect of switching from SnpEff to VEP annotation databases on our pipeline with regard to improving sensitivity and accuracy
- Instituted and led weekly Genome Sciences team training sessions covering use of UNIX and NGS analysis tools to facilitate data analysis

Research Fellow, Dr. Sharon A. Savage Laboratory
Division of Cancer Epidemiology and Genetics, NCI, NIH, Rockville, MD 03/2012-12/2015

- Wrote a set of custom Perl and Bash tools to facilitate whole exome sequencing data analysis of both pedigrees and large populations on the NIH compute cluster
- Used databases and tools including ClinVar, BLAST, GenBank (via the Nucleotide database), RefSeq, and dbSNP in NGS data analysis
- Identified novel causal mutations in patients with inherited cancer predisposition syndromes, resulting in numerous collaborations and 9 publications (5 first or co-first author)

- Authored proposals that were awarded over $60,000 in funding to leverage the DCEG biospecimen collection via innovative NGS analyses

Doctoral Candidate, Dr. Vicki Lundblad Laboratory
Salk Institute for Biological Studies/UC San Diego, La Jolla, CA 09/2005-12/2011

- Led two major projects to explore mechanisms by which telomeres are protected from being misinterpreted by the cell as DNA breaks, resulting in first- and middle-author publications in peer-reviewed journals
- Characterized a novel candidate responsible for single-strand resection of double-strand breaks, a necessary step in repair by homologous recombination
- Developed novel in-house capabilities in molecular biology techniques previously unused in the Lundblad lab; trained other graduate students in these new techniques

TECHNICAL PROFICIENCIES

- NGS data analysis: bedtools, vcftools, bcftools, GATK, IGV
- Workflow management languages: Snakemake
- Containerization: Singularity, Docker
- Jupyter notebooks
- Version control: git, GitHub, GitLab
- Cloud environments for collaboration: Google Colab, Binder
- Package/environment managers: anaconda, pip, homebrew, yum
- Virtual machines: Vagrant
- Programming/scripting languages: Python, R, MATLAB, Perl, VBA, bash, SQL, and C++
- Variant annotation tools: ANNOVAR, SnpEff, VEP, dbNSFP
- *NIX operating systems and Grid Engine- and Slurm-based HPCs
- Genomic/cancer biology data standards: HGVS, SNOMED, GO ontology
- Public genomic databases: gnomad, ESP, ExAC, 1000 Genomes, dbSNP
- Population genetics tools: PLINK

EDUCATION

Ph.D., Biology **University of California, San Diego**, San Diego, CA, 2011
B.S., Molecular and Cellular Biology **Johns Hopkins University**, Baltimore, MD, 2005

SELECTED PUBLICATIONS

Gadalla SH*, **Ballew BJ***, Haagenson M, Spellman S, Hicks B, Alter BP, Zhu B, Zhou W, Yeager M, Wang T, Fleischhauer K, Hsu K, Verneris M, Freedman N, Lee SJ, Savage SA. Germline mutations in marrow failure predisposition genes in patients receiving unrelated donor hematopoietic cell transplant for severe aplastic anemia. ***J Clin Invest.*** Submitted (2016). *authors contributed equally

Burris AM*, **Ballew BJ***, Kentosh JB*, Turner CE, Norton SA, NCI DCEG Cancer Genomics Research Laboratory, NCI DCEG Cancer Sequencing Working Group, Giri N, Alter BP, Nellan A, Gamper C, Hartman KR, Savage SA. Hoyeraal-Hreidarsson syndrome due to *PARN* mutations: fourteen years of followup. ***Pediatr Neurol***. doi: 10.1016/j.pediatrneurol.2015.12.005 (2016). *authors contributed equally

Kocak H*, **Ballew BJ***, Bisht K*, Eggebeen R, Hicks BD, Suman S, O'Neil A, Giri N, NCI DCEG Cancer Genomics Research Laboratory, NCI DCEG Cancer Sequencing Working Group, Maillard I, Alter BP, Keegan CE, Nandakumar J, Savage SA. Hoyeraal-Hreidarsson syndrome caused by a germline mutation in the TEL patch of the telomere protein TPP1. ***Genes Dev***. 28(19):2090-102 (2014). *authors contributed equally

Ballew BJ*, Joseph V*, De S*, Sarek G, Vannier JB, Stracker T, Schrader KA, Small TN, O'Reilly R, Manschreck C, Harlan Fleischut MM, Zhang L, Sullivan J, Stratton K, Yeager M, Jacobs K, Giri N, Alter BP, Boland J, Burdett L, Offit K, Boulton SJ, Savage SA, Petrini JH. A recessive founder mutation in regulator of telomere elongation helicase 1, *RTEL1*, underlies severe immunodeficiency and features of Hoyeraal Hreidarsson syndrome. ***PLoS Genet*** 9(8):e1003695 (2013). *authors contributed equally

****Ballew BJ**, Yeager M, Jacobs K, Giri N, Boland J, Burdett L, Alter BP, Savage SA. Germline mutations of regulator of telomere elongation helicase 1, *RTEL1*, in Dyskeratosis congenita. ***Human Genetics*** 132(4):473-80 (2013). **most frequently cited article of 2013