

Sage M. Wright

Education

- 2021 **M.Sc. Bioinformatics and Genomics**, *Penn State University*, Hershey, PA.
Thesis Title: The Effect of CK2 Inhibition on Heterochromatin in AML
Advisor - Dr. Sinisa Dovat
- 2017 **B.S. Bioinformatics**, *Brigham Young University*, Provo, UT.
Minors: Computer Science, English, Music

Experience

Work

- 2024–now **Senior Scientist, Bioinformatics (Developer)**, *Theiagen Genomics*, Highlands Ranch, CO.
- 2022–2024 **Bioinformatics Scientist**, *Theiagen Genomics*, Highlands Ranch, CO.
- Organize and lead a team of developers as they worked to meet deadlines, resolve issues, and add new features as they worked to create bioinformatic pipelines for public health.
 - Develop and maintain Google Workflows running Google Batch jobs to manage the processing and movement of data through the Google Cloud Platform and various features (BigQuery, Looker, LookerDataStudio, etc.).
 - Develop and maintain bioinformatic pipelines for the analysis of viral, bacterial, and eukaryotic pathogens (see the Theiagen Genomics GitHub repositories).
 - Develop the 'tbp-parser' Python package in collaboration with the California Department of Health to classify *Mycobacterium tuberculosis* mutations that potentially confer antimicrobial resistance.
 - Develop the 'Mercury' Python package (part of the Mercury_Prep_N_Batch workflow) that helps prepare SARS-CoV-2 and mpox genomic data for submission to national and international repositories.
 - Develop the Terra_2_NCBI workflow that prepares *and* submits genomic data for submission to NCBI's BioSample and SRA data repositories.
 - Develop the 'theiavalidate' Python package to compare the results of different bioinformatic workflows which has been use as part of the CLIA validation process in various public health laboratories in the United States.
- 2021–2022 **Bioinformatician**, *Pennsylvania Department of Health Bureau of Laboratories*, Exton, PA.
- Established a cloud-based bioinformatic infrastructure to enable the rapid data analysis of organisms of public health concern to enable appropriate outbreak responses and epidemiological tracing.
 - Analyzed SARS-CoV-2 genomes to ensure proper quality, identify their lineages, and create phylogenetic trees to help track viral evolution.
 - Contributed SARS-CoV-2 genomes and associated metadata to national and international databases, such as NCBI and GISAID.

Research

- 2020–2021 **Graduate Student Researcher**, *Dr. Sinisa Dovatz Lab*, Hershey, PA.
- Analyzed occupancy patterns of heterochromatin histone modifications after inhibition of the protein CK2 using RNA-Seq and ChIP-Seq.
- 2019–2020 **Graduate Student Researcher**, *Dr. James Broach Lab*, Hershey, PA.
- Identified variants associated with disease phenotype in newborn infant from the whole-genome sequencing data of a family.
- Analyzed data produced from optical mapping technology (Bionano).
- Identified genomic variants in HPV-integrated cancers.
- 2017–2019 **Graduate Student Researcher**, *Dr. Feng Yue Lab*, Hershey, PA.
- Identified structural variants in alternative sequencing technologies, such as Hi-C, Bionano optical mapping, and long-read sequencing (ONT).
- Maintained and updated the databases behind the 3D Genome Browser
- Developed a machine learning algorithm to identify TAD regions in 3D genomics data.
- 2015–2017 **Bioinformatics Research Assistant**, *Dr. John Kauwe Lab*, Provo, UT.
- Prepared a pipeline to automate the process of running Natural Selection analyses from aligned genome sequence files.
- Calculated correlation between IL6R and CRP cerebral spinal fluid levels in Alzheimer's Disease patients.
- Determined relationship between rare variants and Alzheimer's Disease status and progression.
- Conducted GWAS analyses for Alzheimer's Disease
- 2016 **Summer Undergraduate Research Intern**, *Dr. Feng Yue Lab*, Hershey, PA.
- Examined the effect of structural and copy number variations in both Osteosarcoma and Parkinson's Disease.

Volunteering

- 2021–now **Member**, *Bioinformatics Pipelines and Visualization Working Group*, Public Health Alliance for Genomic Epidemiology (PHA4GE).
Contributed to the following guidance documents
- Ten Best Practices for Public Health Bioinformatics Pipelines (2024)
- Identifying SARS-CoV-2 Recombinants (2023)
- Bioinformatics Solutions for Mpox Genomic Analysis (2022)

Publications

In reverse chronology

Wadford DA, Baumrind N, Baylis EF, Bell JM, Bouchard EL, Crumpler M, Foote EM, Gilliam S, Glaser CA, Hacker JK, Ledin K, Messenger SL, Morales C, Smith EA., Sevinsky JR, Corbett-Detig RB, DeRisi J, Jacobson K, **the COVIDNet Consortium**. 2023. Implementation of California COVIDNet – a multi-sector collaboration for statewide SARS-CoV-2 genomic surveillance. *Front Public Health* 11. doi:10.3389/fpubh.2023.1249614

Ambrosio FJ, Scribner MR, **Wright SM**, Otieno JR, Doughty EL, Gorzalski A, Siao DD, Killian S, Hua C, Schneider E, Tran M, Varghese V, Libuit K, Pandori M, Sevinsky JR, Hess D. 2023. TheiaEuk: A Species-Agnostic Bioinformatics Workflow for Fungal Genomic Characterization. *Front Public Health* 11. doi:10.3389/fpubh.2023.1198213

Libuit KG, Doughty EL, Otieno JR, Ambrosio F, Kapsak CJ, Smith EA, **Wright SM**, Scribner MR, Petit RA, Mendes CI, Huergo M, Legacki G, Loreth C, Park DJ, Sevinsky JR. 2023. Accelerating bioinformatics implementation in public health. *Microb Genom* 9:001051. doi:10.1099/mgen.0.001051

Smith EA, Libuit KG, Kapsak CJ, Scribner MR, **Wright SM**, Bell J, Morales C, Crumpler M, Messenger S, Hacker JK, Ledin K, Glaser C, Jacobson K, Sevinsky JR, Wadford DA. 2023. Pathogen genomics in public health laboratories: successes, challenges, and lessons learned from California's SARS-CoV-2 Whole-Genome Sequencing Initiative, California COVIDNet. *Microb Genom* 9:001027. doi:10.1099/mgen.0.001027

LaBarge B, Hennessey M, Zhang L, Goldrich D, Chartrand S, Purnell C, **Wright SM**, Goldenberg D, Broach JR. 2022. Human Papillomavirus Integration Strictly Correlates with Global Genome Instability in Head and Neck Cancer. *Mol Cancer Res* 20(9):1420-1428. doi:10.1158/1541-7786.mcr-21-0831

Salameh T, Wang X, Song F, Zhang B, **Wright SM**, Khunsiraksakul C, Yue F. 2020. A supervised learning framework for chromatin loop detection in genome-wide contact maps. *Nat Comm* 11:3428. doi:10.1038/s41467-020-17239-9

Wright SM, Jensen SL, Cockriel KL, Davis B, Tschanz JT, Munger RG, Corcoran CD, Kauwe JSK. 2019. Association study of rs3846662 with Alzheimer's disease in a population-based cohort: the Cache County Study. *Neurobiol Aging* 84:242e1-242e6. doi:10.1016/j.neurobiolaging.2019.03.004

Miller JB, Hippen AA, **Wright SM**, Morris C, Ridge PG. 2017. Human viruses have codon usage biases that match highly expressed proteins in the tissues they infect. *Biomed Genet Genomics* 2(2):1-5. doi:10.15761/BGG.1000134

Wright SM, Carroll C, Walters A, Newell PD, Chaston JM. 2017. Genome sequence of *Leuconostoc citreum* DmW_111, isolated from wild *Drosophila*. *Genome Announc* 5:e00507-17. doi:10.1128/genomeA.00507-17

CV Version: 2024-08-16