

EMPLOYMENT

-
- | | | |
|--|--|------------------------------|
| Software Engineering Intern | UCSC Genomics Institute – Computational Genomics Laboratory | Summer 2018 |
| <ul style="list-style-type: none">Created a Docker container for the Comparative Annotation Toolkit (CAT) and implemented autoscaling of the distributed container by integrating Toil (workflow engine) onto Amazon Web Services.Created a gRPC server for server-side visualization of Hierarchical Alignment (HAL) format on the UCSC genome browser.Ran multiple genomes through the CAT pipeline for gene annotation and implemented unit tests using Travis CI. | | |
| Undergraduate Researcher | UCSC Genomics Institute – Computational Genomics Laboratory | Fall 2017 – Present |
| <ul style="list-style-type: none">Worked on comparative genomics algorithms and maintaining Cactus, a reference-free whole genome multiple alignment based on cactus graphs and improved it by creating a post-filtering algorithm using a pair Hidden Markov Model to prevent likely alignments from being dropped.Created an alignment-free pairwise branch distance predictor of phylogenetic trees to prevent species switching, mismatch, and branch length discrepancies for a pre-check input into Cactus. Eliminated any human or scaling errors in tree building.Generated the whole genome alignment of 19 bumblebee's species and annotated 17 newly assembled bumblebee genomes using Cactus and CAT. (publication pending)Using and modifying the underlying Cactus graph parameters to differentiate the close orthologies of Zinc Finger clusters.Part of the literature review group for Xena, the Cancer Genome Browser specifically on single cell RNA-seq literature. | | |
| Undergraduate Researcher | UCSC – Brooks Lab | Spring 2017 – Present |
| <ul style="list-style-type: none">Identified alternative splice site motifs through detection of modifications in native RNA-seq by overlapping raw nanopore signal level data.Searched for specific A to I modifications by comparing direct RNA-seq to cDNA data from the nanopore and created trained model to search for these sites at the signal level.Created a general genome wide parser for possible modifications in direct RNA-seq sequencing using nanopore by filtering for mismatch frequency.Clustered associated genes and specific modifications of RNA-seq and characterized the modifications by motifs. | | |
| Food Sciences, Intern | USDA Agricultural Research Services (WRRRC) | Summer 2016 |
| <ul style="list-style-type: none">Performed compositional analysis of grape pomace from local Napa Valley wineries.Learned methods on lipid extraction, protein analysis, ash analysis, and moisture collection and compiled the data for future research.Compiled a report on the data and prospects in food science with grape pomaces and the market for such a product based on current supplies of wasted pomace. | | |

EDUCATION

-
- | | |
|---|------------------------------|
| University of California, Santa Cruz | June 2017 – June 2019 |
| <ul style="list-style-type: none">B.S. in Biomolecular Engineering and Bioinformatics (Bioinformatics Concentration) | |
| Relevant Coursework – (G) Graduate Courses | |
| <ul style="list-style-type: none">Computer Science and Engineering: Databases, Data Structures, Discrete Structures, Algorithms, Computer Architecture, Programmable Hardware Accelerators (G)Biology: Cell and Molecular Biology, Development and Physiology, Genetics, Biochemistry, Organic ChemistryBioinformatics: Bioinformatics Models and Algorithms (G), Computational Genomics and Systems Biology (G), Evolutionary Genomics (G)Statistics: Probability Theory, Machine Learning, Applied Machine Learning, Statistical Learning Theory (G), Applied Bayesian Statistics (G)Enrolled Spring 2019: Information Theory (G), Stochastic Modelling in Biology (G), Distributed Systems, Stem Cell Engineering, Algorithmic Foundations of Convex Optimization (G), RNA World, Advanced Computational Genomics and Systems Biology (G) | |

TECHNICAL EXPERIENCE

Projects

- PoE Trader** – <https://github.com/cnk113/POEStash>
Gets up-to-date instances of user's stash data through the public REST API for the use of trading and indexing items.
- Sentiment Analysis of Movie Reviews** – <https://github.com/cnk113/sentiment-analysis-on-movie-reviews>
Did a comparative analysis of a linear kernel Support Vector Machine and bidirectional RNN for classification of movie reviews. Created custom lemmatizer, tokenizer, and preprocessor.

ADDITIONAL EXPERIENCE AND AWARDS

- American Chemical Society SEED Scholarship:** Awarded a \$5,000 scholarship based on research conducted at the USDA
- Korean Ancestry Grant:** \$2,500 scholarship for academic excellence, GPA 3.5+

Skills

- Java (proficient), C++ (familiar), JavaScript (familiar), Python (advanced), C (familiar), Assembly (familiar), R (familiar), SQL, AWS, Azure, Unix, Git, Docker, Bioinformatics Tools (minimap2, Samtools, STAR, etc.), ML tools (sklearn, Keras, NLTK, TensorFlow)