SIDHANT PUNTAMBEKAR

Computer Scientist | Computational Biologist

sidhantnp@yahoo.com
720-250-8098
https://www.linkedin.com/in/sidhant-puntambekar/
https://github.com/SidhantPuntambekar

Education

Bachelor of Arts in Computer Science, Minors in Computational Biology and Molecular Biology
University of Colorado Boulder*,* Boulder*,* CO*, August 2019 - May 2023*

- **Honors:** Graduated with distinction honors. **Awards:** Discovery Learning Award. **GPA*:* 3.980 Major | 3.965 Cumulative**

Work Experience

Bioinformatician, Dr. Matthew Sampson Lab for Kidney Genomics, *Boston Children's Hospital,* Boston, MA, *June 2024 - present*
Affiliated Kidney Disease Initiative Researcher, *Broad Institute of MIT and Harvard,* Boston, MA, *June 2024 - present*

- Maintain computational pipeline for eQTL discovery in immune cells from 211 steroid sensitive nephrotic syndrome patients leveraging bulk RNA-seq, WGS, PCA, and bayesian prior analysis for eGene fine mapping culminating in a research paper draft.
- Develop high-level software design and code for the Biobank to Illuminate the Genomics of Kidney Diseases (BIGKiDs) study at the Sampson Lab through an interactive R Shiny application.
- Mentor pediatric nephrology medical research fellows in creating code implementations for nephrotic syndrome research projects adhering to software development best practices as well as rigorous unit and regression testing.

Bioinformatics Analyst, Bioinformatics and Software Development Team
Bionano Genomics**,** San Diego, CA, *January 2023 - May 2024*

- Developed pipelines, algorithms, and validations for detection of copy number variations and structural variants in cytogenetics data from optical genome mapping (OGM) and next generation sequencing (NGS) methods as part of an agile development team.
- Spearheaded continuous integration testing and automated unit testing efforts for Bionano Solve codebases and Singularity builds including 22 repositories composing 5 distinct analysis pipelines.
- Curated structural variant relational control databases in Perl comprising 450+ OGM samples for production releases of Bionano Solve 3.8 and 3.8.1, Access 1.8, and VIA 7.0 software suites.

Computational Biology/Data Science Intern, Multi-Cancer Early Detection (MCED) Team
Exact Sciences Corporation, Boston, MA, *May 2022 - August 2022*

- Conducted statistical analysis in Python and R resulting in a greater than 95% accuracy improvement of genomic mutational variant biomarker signal in CancerSEEK, a developmental liquid biopsy diagnostic for the early detection of stage I and II cancers.
- Analyzed effects of unique molecular identifier (UMI) contamination on 40 genomic primer targets in the SafeR-SeqS duplex sequencing assay to improve sensitivity and specificity of a rare variant machine learning based logistic regression classifier.
- Leveraged Apache Spark and Microsoft Azure Databricks to quantify UMI contamination from over 430,000+ NGS reads.

Biomedical Informatics Research Intern, Dr. Shamil Sunyaev Computational Genomics Lab
Harvard Medical School**,** Boston, MA, *May 2021 - May 2022*

- Analyzed the Broad Institute's Genome Aggregation Database (gnomAD) dataset to better quantify and understand the mutational constraint on missense and loss of function genomic variants using a fine-scale mutational map.
- Modeled baseline mutational spectrum using machine learning techniques (logistic regressions) on a training set of over 125,000 synonymous genome and exome sequences leveraging scikit-learn, pandas, and dask.
- Built a logistic regression classifier using 10 independent covariates to predict tolerances of genes that are more likely to succumb to debilitating loss of function mutations.

Bioinformatics Research Intern, RNA Bioscience Initiative Informatics Fellows Group
University of Colorado Anschutz Medical School, Denver*,* CO*, May 2020 - May 2021*

- Authored (primary authorship) research paper published in Public Library of Sciences Biology journal (May, 2021) regarding the problem of single cell RNA-sequencing metadata submission in gene expression databases (NCBI GEO).
- Developed RShiny web application for the open source 'ClustifyR' R library to facilitate analysis, clustering, and cell identification of scRNA-seq data by 8 RNA Bioscience Initiative wet lab scientists in collaboration with the informatics fellows group.
- Created a meta-analysis of 15 scRNA-seq data sets from NCBI Gene Expression Omnibus and Tabula Muris (mouse scRNA-seq atlas) utilizing Seurat, principal component analysis, graph based clustering, and uniform manifold approximation and projection.

Skills & Technologies

Languages: Python, Java, C, C++, R, JavaScript, TypeScript, HTML5, CSS, SQL, MATLAB, Scala, Perl, Bash
Testing Tools: Mocha, Chai, Jest, JUnit, Pytest
Continuous Integration Tools: Jenkins, TravisCI, Bitbucket pipelines
Production Tools: AWS (S3, EC2), Azure, Azure Databricks, Apache Spark, Git, Heroku, Docker, Singularity, npm, Yarn
JavaScript Technologies: React, Node.js, Express.js, Passport.js, Svelte, Bootstrap, TailwindCSS
Databases: MongoDB, MySQL, PostgreSQL
High Performance Computing Tools: Sun Grid Engine, Slurm