Matthew Sooknah

1552 East Gate Way, Pleasanton CA 94566
401-588-2644 — mattsooknah@gmail.com — mattsooknah.github.io

Research Interests

- DNA and RNA sequence analysis
- Visualization of large biological data sets
- New sequencing technologies and assays
- Applying machine learning and graph / network algorithms to biological problems

Education

Massachusetts Institute of Technology **Cambridge, MA**
S.B. Physics, GPA 4.9/5.0 *2009 – 2013*

- Relevant coursework: Computational biology (graduate-level), combinatorial optimization (graduate-level), algorithms, linear algebra, differential equations, statistical mechanics (including probability and statistics)

Research Experience

10X Genomics **Pleasanton, CA**
Scientist, Computational Biology Group *January 2016 – present*

- Developed more efficient algorithms for aggregation and analysis of large-scale single cell gene expression data.
- Prototyped new methods leveraging molecular barcodes to improve transcriptome mapping rates in RNA sequencing data.
- Contributed to design and analysis of a custom exome bait set that uses 10X linked reads to improve gene phasing.
- Contributed to the development of the Long Ranger and Cell Ranger pipelines for analyzing 10X data.

The Broad Institute of MIT and Harvard **Cambridge, MA**
Software Engineer, Data Sciences & Data Engineering Group *May 2014 – December 2015*

- Developed pipelines to process petabyte-scale sequencing data produced by the Broad Genomics Platform.
- Contributed to development and support of the Picard and HTSJDK open source toolkits for analyzing sequencing data.
- Wrote tool to measure incidence of sequencing errors caused by oxidative damage to DNA during preparation of short-read sequencing libraries.
- Helped implement a pipeline for rapid processing and QC of sequencing runs.
- Contributed to FireCloud, a cloud-based platform for analyzing TCGA data within user-defined workspaces.
- Developed methods for analyzing gene expression and pathway activity from bulk RNA-seq and TF-seq (a novel assay) to gain insight into immune cell behavior.

Nabsys **Providence, RI**
Associate Scientist, Algorithms Group *May – August 2012, June 2013 – May 2014*

- Prototyped algorithms for assembly of genomic maps from Nabsys data, which reports the positions of short, predefined motifs on long DNA input molecules using recognition probes.
- Wrote tools for visualization of Nabsys data and identification of structural errors in DNA map assembly.

- Implemented a signal processing pipeline for extracting information about molecules and recognition probes from an electronic readout.
- Assisted in modeling how molecules behave under the influence of many forces (electromotive force, viscous drag, brownian motion, etc) within the Nabsys microfluidic system, and the resulting sources of error.

MIT SETG Lab / Massachusetts General Hospital **Boston, MA**
Undergraduate Researcher *February – May 2013*

- Worked in the lab of Professors Gary Ruvkun and Maria Zuber on SETG (the Search for Extra-Terrestrial Genomes).
- Helped characterize bacterial composition of samples from extreme environments (e.g. acid lakes) using a variety of published metagenomics tools,
- The work was part of an effort to define approaches for collection and analysis of potential biological samples on a future Mars mission.

MIT Kavli Institute for Astrophysics **Cambridge, MA**
Undergraduate Researcher *February – August 2011*

- Worked in the lab of Professor Paul Schechter on gravitational lensing.
- Analyzed how methods for characterizing weak gravitational lensing perform on large data sets of galaxies, such as the Sloan Digital Sky Survey, as well as simulated data.
- Performed image processing and feature extraction, implemented lensing models, and characterized sources of error.

Publications

- O'Connell DJ, Kolde R, **Sooknah M**, et al. 2016. Simultaneous Pathway Activity Inference and Gene Expression Analysis Using RNA Sequencing. Cell Systems 2016; 2(5): 323–334. PMID 27211859.

Presentations

- "Mapping, processing, and duplicate marking with Picard tools." BroadE Workshop on GATK Best Practices. Broad Institute, Cambridge, MA. March 2015.