

# Matt Sooknah

Scientist, 10X Genomics  
mattsooknah@gmail.com  
mattsooknah.github.io

1552 East Gate Way  
Pleasanton, CA 94566

## 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

### S.B. Physics, Massachusetts Institute of Technology (2009-2013)

- GPA: 4.9/5.0
- Relevant classes: Computational biology (graduate-level), combinatorial optimization (graduate-level), algorithms, linear algebra, statistical mechanics (including statistical methods)

## Research Experience

### 10X Genomics (Jan 2016 - Present)

*Scientist, Computational Biology Group*

- 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.

### Broad Institute of MIT and Harvard (May 2014 - Dec 2015)

*Software Engineer, Data Sciences & Data Engineering Group*

- 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 (novel assay) to gain insight into immune cell behavior.

## **Nabsys (summer 2012 internship, June 2013 - May 2014)**

*Associate Scientist, Algorithms Group*

- 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 (February-May 2013)**

*Undergraduate Research Assistant - Professors Gary Ruvkun and Maria Zuber*

- Helped characterize bacterial composition of samples from extreme environments (e.g. acid lakes) using a variety of published metagenomics tools.

## **MIT Kavli Institute for Astrophysics (February-August 2011)**

*Undergraduate Research Assistant - Professor Paul Schechter*

- 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.

## **Papers**

- O’Connell DJ, Kolde R, **Sooknah M**, et al. 2016. Simultaneous Pathway Activity Inference and Gene Expression Analysis Using RNA Sequencing. Cell Systems 2, 323–334. <http://dx.doi.org/10.1016/j.cels.2016.04.011>

## **Presentations**

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