Stanford

Somalee Datta

Director of Research IT, Technology & Digital Solutions

Bio

BIO

I am a physicist by training and a biotechnologist by profession. I believe that with the explosion of data in healthcare and with new methods to analyze such large amounts of data, we will see massive changes in how human diseases are addressed via novel drugs, large scale genomics, wearable sensors, and software to tie it all together. I want to drive part of this revolution.

Prior to joining Stanford in 2012, I spent a dozen years at various biotechs in the Bay Area. This includes experiences as technology lead at Life Technologies (now Thermo Fisher) and founding team member of Verseon, a drug discovery company. Along the way, I have had fantastic opportunities to work alongside some of the smartest people in the field, learn from some of the most brilliant minds of our times, solve some fundamental technological problems, and delivered business impact.

CURRENT ROLE AT STANFORD

I am currently the Director of Research IT at School Medicine. Research IT is a critical part of Stanford's Precision Health Strategy and exists to supply infrastructure, tools, and services used by researchers, patients/participants, and clinicians to collect and combine data to make discoveries and to improve human health and wellness. Our team builds and maintains STARR (STAnford medicine Research data Repository), Stanford REDCap, CHOIR and mHealth platforms and builds custom applications to streamline hundreds of studies.

I joined Stanford in Oct 2012 as the Director of Bioinformatics at Stanford Center for Genomics and Personalized Medicine (SCGPM). My responsibility at the Center was to develop and lead the bioinformatics team and establish a genomics data analysis facility. Currently, SCGPM bioinformatics team is comprised of a dozen scientists and software engineers. The team has a wide range of skill sets including omics, computational biology, machine learning, software engineering, data management, Databases, Visualization, High Performance Computing, IT, and Cloud DevOps. The team is currently supporting several large scale research and clinical programs at Stanford including prestigious consortium efforts and inter-disciplinary collaborations. The team also supports Genetics Bioinformatics Service Center (2013-), a facility that provides best-in-class high performance computational systems, scalable Cloud computing and cutting edge bioinformatics services for the Stanford community.

EDUCATION AND CERTIFICATIONS

- PhD, Boston University, MA, USA , Computational Physics (non-equilibrium statistical mechanics) (2000)
- MSc, Indian Institute of Technology, Madras (aka Chennai), India , Physics (stochastic systems) (1994)
- BSc, Jadavpur University, Calcutta (aka Kolkata), India , Physics (1992)

LINKS

- Research IT, Technology & Digital Solutions: http://med.stanford.edu/researchit.html

- LinkedIn: https://www.linkedin.com/in/somaleedatta

Publications

PUBLICATIONS

- **Benchmarking workflows to assess performance and suitability of germline variant calling pipelines in clinical diagnostic assays.** *BMC bioinformatics*
  Krishnan, V., Utiramerur, S., Ng, Z., Datta, S., Snyder, M. P., Ashley, E. A.
  2021; 22 (1): 85

- **Large-Scale Assessment of a Smartwatch to Identify Atrial Fibrillation.** *The New England journal of medicine*
  Perez, M. V., Mahaffey, K. W., Hedlin, H., Rumsfeld, J. S., Garcia, A., Ferris, T., Balasubramanian, V., Russo, A. M., Rajmane, A., Cheung, L., Hung, G., Lee, J., Kowey, et al
  2019; 381 (20): 1909–17

- **SciReader: A Cloud-based Recommender System for Biomedical Literature**
  Desai, P., Telis, N., Lehmann, B., Bettinger, K., Pritchard, J. K., Datta, S.
  bioRxiv.
  2018

- **Digital Health: Tracking Physiomes and Activity Using Wearable Biosensors Reveals Useful Health-Related Information.** *PLoS biology*
  Li, X., Dunn, J., Salins, D., Zhou, G., Zhou, W., Schüssler-Fiorenza Rose, S. M., Perelman, D., Colbert, E., Runge, R., Rego, S., Sonecha, R., Datta, S., McLaughlin, et al
  2017; 15 (1)

- **Cloud-based interactive analytics for terabytes of genomic variants data.** *Bioinformatics (Oxford, England)*
  Pan, C. n., McInnes, G. n., Deflaux, N. n., Snyder, M. n., Bingham, J. n., Datta, S. n., Tsao, P. S.
  2017; 33 (23): 3709–15

- **Secure cloud computing for genomic data** *Nature Biotechnology*
  Somalee, D., Keith, B., Michael, S.
  2016; 34 (6): 588-91

- **Sequence to Medical Phenotypes: A Framework for Interpretation of Human Whole Genome DNA Sequence Data** *PLOS GENETICS*
  Dewey, F. E., Grove, M. E., Priest, J. R., Waggott, D., Batra, P., Miller, C. L., Wheeler, M., Zia, A., Pan, C., Karzcewski, K. J., Miyake, C., Whirl-Carrillo, M., Klein, et al
  2015; 11 (10)

- **The Integrative Human Microbiome Project: Dynamic Analysis of Microbiome-Host Omics Profiles during Periods of Human Health and Disease** *CELL HOST & MICROBE*
  Proctor, L. M.
  2014; 16 (3): 276-289