Diya Das, PhD

diyadas@alumni.princeton.edu | linkedin.com/in/diyadas | web: diyadas.github.io

Experience

Development Sciences Informatics, Genentech
Senior Informatics Analyst — July 2021 – Present
Informatics Analyst III — May 2020 – June 2021
Informatics Analyst II — June 2019 – April 2020

- Data Planning (Lead, March 2020 – Present)
  - Collaborate with Biomarker Operations to obtain biomarker data and metadata from CROs and academic collaborators in standardized formats.
  - Write data transfer specifications and collaborate with IT partners to set up file transfer protocols.
  - Oversee the work of 6-7 consultants and 1 FTE (converted from consultant in November 2020).
- Data Sharing & Anonymization (Lead, March 2020 – Present)
  - Anonymize and export data in support of Roche & Genentech publications and research collaborations.
  - Collaborate with internal patient data sharing experts, legal counsel and contracts groups to ensure data is shared according to best practices for protecting patient privacy while advancing research interests.
  - Share knowledge of data sharing scenarios and associated guidance with cross-Roche working groups.
  - Oversee the work of 2 consultants (since April 2021).
- Department representative, DevSci Manager Effectiveness Committee — April 2021 – Present
- Intern mentor — January 2020 – December 2020
  - Cohort mentor for department's first six-month intern cycle (met every 3 weeks with all interns).
  - Mentored 4 interns (1 from January - May 2020, 3 from July - December 2020).
- Collaborated with Sivan Cohen's team in BAS to analyze single-cell data to assess immunogenicity.
- Devised and updated an onboarding checklist for new members of the data management team.
- Set up a monthly meeting series and chatroom with our IT partners to facilitate operational activities.

Lab of John Ngai, UC Berkeley
Postdoctoral Researcher — June 2018 – May 2019
Graduate Student Researcher — May 2013 – May 2018

- Developed analysis pipelines for Fluidigm and 10X Genomics single-cell RNA-sequencing data.
- Identified when cells choose between neuronal and non-neuronal fates by applying resampling-based ensemble clustering and lineage trajectory algorithms developed with statisticians and computer scientists.
- Explored gene regulation by chromatin accessibility in stem cells using ATAC-seq.
- Wrote scripts to retrieve sequencing data and accompanying experimental metadata from NCBI databases.

Education

PhD in Molecular & Cell Biology, University of California, Berkeley — August 2012 – May 2018
AB in Molecular Biology, Princeton University — September 2008 – June 2012
Certificate in Neuroscience, Quantitative and Computational Neuroscience honors track

Publications and Preprints (* = co-first author)

- A. Arefolov, L. Adam, S. Brown, Y. Budovskaya, C. Chen, **D. Das** et al. (2021). Implementation of The FAIR Data Principles for Exploratory Biomarker Data from Clinical Trials. Data Intelligence: https://doi.org/10.1162/dint_a_00106
- D.H. Brann, T. Tsukahara, C. Weinreb, M. Lipovsek, K. Van den Berge, B. Gong, R. Chance, I.C. Macaulay, H.J. Chou, R.B. Fletcher, **D. Das** et al. (2020). Non-neuronal expression of SARS-CoV-2 entry genes in the olfactory system suggests mechanisms underlying COVID-19-associated anosmia. Science Advances *6*, eabc5801.
- **D. Das**, R.B. Fletcher and J. Ngai. Cellular Mechanisms of Epithelial Stem Cell Self-Renewal and Differentiation During Homeostasis and Repair. (2020). *Invited review.* WIREs Developmental Biology *9*, e361.
- D. Sholler*, **D. Das***, F. Hoces de la Guardia* et al. (2019). Best Practices for Managing Turnover in Data Science Groups, Teams, and Labs. SocArXiv: https://doi.org/10.31235/osf.io/wsxru.
- R.S. Geiger, O. DeMasi, A. Culich, A. Zoglauer, **D. Das** et al. (2019). Best Practices for Fostering Diversity and Inclusion in Data Science. SocArXiv: https://doi.org/10.31235/osf.io/8gsjz.
- D. Risso, L. Purvis, R. Fletcher, **D. Das** et al. (2018). clusterExperiment and RSEC: A Bioconductor package and framework for clustering of single-cell and other large gene expression datasets. PLOS Computational Biology *14*, e1006378.
- R.B. Fletcher, **D. Das** and J. Ngai. (2018). Creating Lineage Trajectory Maps Via Integration of Single-Cell RNA-Sequencing and Lineage Tracing. BioEssays *40*, 1800056.
- K. Street, D. Risso, R.B. Fletcher, **D. Das** et al. (2018). Slingshot: Cell lineage and pseudotime inference for single-cell transcriptomics. BMC Genomics *19*, 477.
- R.S. Geiger, C. Mazel-Cabasse, C. Cullens, L. Norén, B. Fiore-Gartland, **D. Das** and H. Brady. (2018). Career Paths and Prospects in Academic Data Science: Report of the Moore-Sloan Data Science Environments Survey. SocArXiv: https://doi.org/10.31235/osf.io/xe823.
- L. Gadye*, **D. Das***, M.A. Sanchez* et al. (2017). Injury Activates Transient Olfactory Stem Cell States with Diverse Lineage Capacities. Cell Stem Cell *21*, 775-790.e9.
- R.B. Fletcher*, **D. Das*** et al. (2017). Deconstructing Olfactory Stem Cell Trajectories at Single-Cell Resolution. Cell Stem Cell *20*, 817-830.e8.

Informal Writing

- **D. Das**. Lessons learned from the Community Discussion on Burnout. *Carpentries Blog*. September 26, 2019. https://carpentries.org/blog/2019/09/community-burnout/.
- **D. Das**. Unlocking the secrets of stem cell renewal. *Berkeley Institute for Data Science: Data Science Insights.* May 22, 2019. https://bids.berkeley.edu/news/unlocking-secrets-stem-cell-renewal.

Fellowships

- Moore-Sloan Data Science Fellow, Berkeley Institute for Data Science August 2017 – May 2019
- Elizabeth Roboz Einstein Fellow in Neurosciences & Human Development January 2015 – May 2015
- California Institute for Regenerative Medicine Predoctoral Fellow January 2015 – December 2015

Talks, Tutorials and Presentations

- Panelist, Data Science Careers, for UC Berkeley graduate course: MCB 295: Careers for Life Science PhDs. March 29, 2021.
- **D. Das** et al. Data Privacy Meets Data Reuse: Creating a Process to Share DevSci Data Externally. 2020 gRED Informatics Symposium, Virtual Poster & Panel. November 9 and 13, 2020.
- **D. Das** et al. Best-Laid Plans: How a Robust Data Planning Strategy Helps Us Make Data Accessible for DevSci Scientists. 2020 gRED Informatics Symposium, Virtual Poster & Panel. November 9 and 13, 2020.
- **D. Das** and N. Gupta. Enabling science through data management and engineering. Northeastern University DATA Club, Boston, Massachusetts. February 13, 2020.
- **D. Das** and D. Huppenkothen. Making the Most of Your Data Science Institute. 2019 Moore-Sloan Data Science Environments Summit, Santa Fe, New Mexico. November 7, 2019.
- **D. Das**. Enabling science through better data management. 2019 Moore-Sloan Data Science Environments Summit, Santa Fe, New Mexico. November 6, 2019.
- **D. Das**. Unraveling Adult Tissue Regeneration. 2018 Moore-Sloan Data Science Environments Summit, Park City, Utah. October 10, 2018.
- **D. Das**, K. Street and D. Risso. Analysis of single-cell RNA-seq data: Dimensionality reduction, clustering, and lineage inference. BioC 2018, Toronto, Ontario. July 27, 2018.
- **D. Das**. Injury Activates Transient Olfactory Stem Cell States with Diverse Lineage Capacities. UC Berkeley Developmental & Regenerative Biology Retreat. November 14, 2017.
- **D. Das**. Deconstructing Olfactory Stem Cell Trajectories at Single-Cell Resolution. UC Berkeley Developmental & Regenerative Biology Retreat. January 9, 2017.
- Various tutorials on shell, GitHub, R and genomic data analysis. UC Berkeley. 2017-2018.

Leadership

Berkeley Institute for Data Science

Best Practices and Meta-Research Working Group	October 2018 – May 2019
Career Paths and Alternative Metrics Working Group	October 2017 – May 2019
Executive Committee, Fellow Representative	June 2018 – November 2018
Strategic Planning Committee	June 2018 – August 2018

Beyond Academia — May 2016 – March 2018

- Planned two 2-day annual conferences for 300+ peers on career options outside academia, featuring 100+ speakers, with team of ~20 graduate students and postdocs.

Co-Director and Development Lead — January 2017 – January 2018

- Organized recruiting and held 1-1 informational meetings with 10 prospective members; 6 joined.
- Managed partnerships with campus units and planned on-campus recruiting event for employers.

Logistics, Speakers and Development Committees — May 2016 – March 2017

- Developed conference schedule for 32 workshops/panels based on 100+ speaker availabilities.
- Organized and coordinated speakers for four panels.
- Redefined advisory board responsibilities and selected new advisors to fulfill needed expertise.

CDIPS Data Science Workshop, **Co-Director**	January 2017 – August 2017
MCB Graduate Student & Alumni Association, **Student Co-President**	June 2015 – May 2017
MCB Graduate Affairs Committee, **Student Representative**	August 2013 – May 2015
Student Health Advisory Committee, **Grad Student Representative**	September 2012 – April 2016
Expanding Your Horizons at Berkeley, **Finance Agent and Signatory**	September 2012 – April 2015