# Prateek Arora, PhD


| | |
|---|---|
| Mobile | +41 782136444 |
| E-mail | prateek@prateekarora.net/ prateek.arora@unibe.ch |
| Website | https://www.prateekarora.net |
| LinkedIn | www.linkedin.com/in/pratarora/ |
| GitHub | https://github.com/pratarora |
| ORCiD | https://orcid.org/0000-0003-0822-9240 |
| Address | Solothurnstrasse 16, 3322, Urtenen-Schönbühl, Switzerland |

## Bioinformatics Research Associate

### SUMMARY


AI-Native Bioinformatician & Data Scientist with 6+ years of experience transforming messy, high-throughput experimental data into high-quality structured datasets. Expert in building reproducible pipelines (Snakemake/Nextflow/Docker) and custom tools (LipidLocator) that bridge computational workflows with wet-lab needs. Fluent in agentic coding (Claude Code, Copilot) to rapidly build reusable functions, optimize token strategies, and ship robust data tools. Combines deep biological grounding in multi-omics with statistical rigor to eliminate batch effects and instrument drift.


### KEY SKILLS

**High-Throughput Analysis:** scRNA-seq (10x, Parse, Nanopore), Spatial Transcriptomics (Visium, Nanostring), Proteomics, Lipidomics, ATAC-seq, Assay Data Quality Control & Statistical Modeling
**Pipeline Development and AI:** Snakemake, Nextflow, Docker, Singularity, Git, FAIR Data Standards, HPC (SLURM), Agentic Development (Claude Code, GitHub Copilot, Local LLMs, Token Optimization, Reusable Functional Architecture).
**Programming & Visualization:** R, Python, UNIX/Bash, Shiny App Development (Full-stack).
**Consultancy:** Experimental design, Statistical consulting, Data visualization, Manuscript preparation.

### EXPERIENCE

**RESEARCH ASSOCIATE - BIOINFORMATICS** **July 2023 - Present**

**University of Bern,** Bern, Switzerland

*Acting as a de facto core bioinformatician for two groups in department with collaborations across more than 4 groups.*

- **Single-Cell & Spatial Innovation:** Led the analysis of >30 scRNA-seq datasets (10x, Parse) and pioneered the department's first Oxford Nanopore long-read snRNA-seq workflow. Expertly handled Spatial Transcriptomics (Nanostring, 10x Visium) projects to study in situ vaccination in cancer, utilizing SingleR and ClusterProfiler for niche-specific pathway analysis.
- **Tool Development:** Conceptualized and built 'LipidLocator', an open-source R Shiny tool for spatial lipidomics. Optimized workflows to reduce project turnaround time by ~98% (from 3 months to 3 days).
- **Multi-omics Integration:** Successfully integrated ATAC-seq with RNA-seq, and Proteomics with Transcriptomics to identify common regulatory pathways
- **Reproducibility Standards & AI Native Workflows:** Enforced strict reproducibility standards using Docker and Snakemake across all computational pipelines. Leveraged agentic workflows (Claude Code/Copilot) to scale capacity, building a customized repository of reusable functions and skills that allowed simultaneous execution of multiple complex projects while ensuring highly standardized, error-checked, and reproducible analytical results across all datasets.

- **Experimental Coordination & Design:** Served as primary liaison to the Sequencing Core Facility, collaborating closely with wet-lab scientists to optimize experimental design, select appropriate chemistries, and troubleshoot raw instrument data anomalies at the source to guarantee benchmark-grade data quality
- **Training & Mentorship:** Introduced and trained PhD students in RNA-seq analysis, pathway enrichment (ClusterProfiler), and functional interpretation, empowering them to perform independent data exploration.
- **Infrastructure Management:** Established and currently maintain the group's high-performance Linux workstation; ensured 100% environment stability.
- **Machine Learning:** Trained and improved AI pipeline to detect injury and healthy heart in AFOG stained hearts based on Unet and TensorFlow.

**POSTDOCTORAL FELLOW - BIOINFORMATICS** **January 2021 – June 2023**

**University of Bern,** Bern, Switzerland

- **Reproducible Research:** Built robust pipelines (Snakemake) for ATAC-seq studies (WT1 transcription factor), utilizing MACS2, DiffBind, and MEME suite.
- **Algorithm Development:** Developed a Ligand-Receptor Network pipeline using bulk RNA-seq and Network theory, directly facilitating two PhD projects.

**SENIOR BIOINFORMATICIAN** **June 2019 – June 2020**

**Elucidata Data Consulting Pvt Ltd**, New Delhi, India

- **Pharma Consulting:** Led a team developing scalable drug target discovery pipelines on AWS for pharmaceutical partners; identified 20+ potential targets.
- **Project Management:** Served as Line Manager for a team of two; managed timelines, validated scientific products, and aligned deliverables with client OKRs.

## CERTIFICATIONS/TRAININGS

- Diving into deep learning - theory and applications with PyTorch (SIB Swiss Institute of Bioinformatics)
- Introduction to Machine Learning with Python (SIB Swiss Institute of Bioinformatics)
- Long-read sequence analysis (SIB Swiss Institute of Bioinformatics)

## PUBLICATIONS/CONFERENCES

- **Organized Conference from scratch** - Exploring regeneration through Omics, Funding acquisition of CHF10000, Inviting speakers, Organizing logistics etc. June 2025
- **Oral Presentation**s - Shiny Conf 2024, Swiss Zebrafish meeting 2023
- **Selected publications**:
  - Arora, P., et al, Bioinformatics Advances, 2026 https://doi.org/10.1093/bioadv/vbag012 (Corresponding author)
  - Garcia-Poyatos, C., Arora, P., et al, Dev Cell, 2024- https://doi.org/10.1016/j.devcel.2024.04.012
  - Arora,P. et al, eLife, 2020 - https://doi.org/10.7554/eLife.49064
- **Total Publications:** 11
- **Publications in Pipeline**: 5

## EDUCATION

**PhD + MSc (Cellular and Molecular Biology)** **2011 - 2020**

Tata Institute of Fundamental Research, Mumbai, India

## LANGUAGES

English - Proficient User (C1/C2 level) | German - (B1 – Intermediate – Give tutorials in German) | Hindi - Mother Tongue