

Serghei Mangul, PhD

Statement of Teaching

I am passionate about teaching computational skills in the interdisciplinary area of computational biology and bioinformatics. I have an extensive background in developing teaching traditional classroom courses for undergraduate and graduate students, as well as hands-on intensive workshops for life sciences and medical researchers. In addition to my teaching experience, I am passionate about mentoring. I believe that integrating undergraduate students directly into genomics research benefits both students and the laboratory members. I have mentored over 20 undergraduate students, three of whom have since joined the Bioinformatics Interdepartmental Ph.D. Program at UCLA.

As a Fellow at the Collaboratory of the Institute for Quantitative and Computational Biosciences (QCBio) at the University of California, Los Angeles (UCLA), I have been teaching computational techniques to life sciences and medical researchers, many of whom lack formal training in use of the command line. I have developed an effective series of three-day workshops that are successful in training students with no prior computational background to use the command line for specific analytical tasks. The goal of these workshops is for life sciences and medical researchers to acquire just enough knowledge and skills to independently use small, yet powerful, commands for rapid exploration and modification of their data. While existing workshops charge a fee and restrict accessibility of materials, I have freely distributed the materials and video recordings of my workshops so anybody can take the course online (<https://github.com/QCB-Collaboratory/W1.UNIX.command.line/wiki>).

Over a span of six years, I taught UNIX workshops to over 500 students, including undergraduates, graduate students, postdoctoral scholars, and faculty from biological and medical disciplines. Qualitative feedback has been overwhelmingly positive; after nine hours of instruction, most participants report that they are able to effectively use the command line to manage and analyze their data. Many participants report mastering fundamental skills, such as directly entering commands line-by-line into a terminal — without the familiar aid of a graphical user interface (GUI).

Teaching these workshops created an opportunity for me to think about the digital divide that exists in contemporary biology. This divide presents a barrier for life sciences and medical researchers seeking to transition from using a GUI (e.g., Microsoft Excel) to UNIX command line. Recently, I published a paper that describes this digital divide and presents a series of principles of teaching the computational techniques for life science and medical researchers. Principles I formulated include, first, minimization of jargon and discipline-specific technical terminology. When unavoidable, I introduce terminology with a clear definition and explanation of the term's context and application. Second, it is necessary to present concepts at a slow and incremental pace. First-time learners often advance in the workshop at different paces. I regularly pause the course to walk around the class and provide one-on-one tutoring. These individual sessions provide an opportunity for students to ask questions that they might otherwise not ask in front of the class. In my workshops, new concepts are introduced stepwise, building upon previous concepts. Thus, I view such interruptions as a way to guarantee that

students acquire the necessary skills before moving on to the next unit. A detailed account of these principles and philosophy of teaching UNIX is presented in my recent paper. I have also shared this work as a highlighted talk at the Education Workshop as part of the *2018 Intelligent Systems for Molecular Biology (ISMB 2018)* conference.

- Mangul, Serghei, et al. "Addressing the digital divide in contemporary biology: Lessons from teaching UNIX.", *Trends in Biotechnology*, <http://dx.doi.org/10.1016/j.tibtech.2017.06.007>

In the traditional classroom setting, I have taught upper-division computational biology and genetics classes at the University of California Los Angeles, including *"Introduction to Computational and Systems Biology"* and *"Advanced Human Genetics Course"*. I am also interested in creating platforms for mentoring undergraduate students in bioinformatics and computational biology. I have served as a bioinformatics advisor for the Bruins-In-Genomics (B.I.G.) Summer Research Program from 2016 to 2018. The B.I.G. Summer Research Program is an eight-week, full-time immersion program for undergraduates interested in doing research in bioinformatics and computational biology. Insufficient training, time, and incentive are among the most substantial barriers preventing the implementation or expansion of research opportunities for undergraduates. However, I believe that involving undergraduates in lab research as an active learning component can actually improve productivity and strengthen research outcomes for senior lab members.

I have demonstrated that incorporating undergraduates in genomics research benefits graduate students, post-doctoral scholars, and faculty. I believe that the analysis of genomic data has specific elements that are uniquely well-suited for successful involvement of undergraduates. Today's sequencing methods produce genomic datasets at an unprecedented scale in terms of size and complexity. This "data explosion" creates several unique challenges and opportunities that are ideal for training undergraduates and leveraging student participation in research. For undergraduates who are primarily involved in the life sciences, participating in computational genomic research can be a transformational experience in interdisciplinary teamwork that increasingly characterizes modern life sciences research. I have developed a framework for involving undergraduates in genomics research and overcoming existing research-and-teaching barriers. Those principles were recently published in *Nature Biotechnology*:

- S. Mangul, L. S. Martin, and E. Eskin. "Involving undergraduates in genomics research to narrow the education–research gap" *Nature Biotechnology* 36.4 (2018): 369.

Some of those principles include identifying particular "low-level" tasks that may take up to a week for an undergraduate to complete. For example, many projects require preparation of computational software tools and pipelines to analyze high throughput data. Installing and running the third-party software tools is often an extremely complicated and time-consuming process—specifically when detailed documentation is unavailable for the software tool. Second, I encourage students to "outsource" foundational education needs through workshops, online resources, and review articles. For example, students who have not previously been exposed to command-line systems are encouraged

to enroll in or self-tutor using in-person or online UNIX workshops, which I have developed as a Fellow at QCB Collaboratory.

I observe several substantial benefits that UREs bring to our computational genomics lab and our broader fields of study. Undergraduate students have substantially contributed to at least five of my recently published papers, some of which have been accepted in high-rank journals, while others are still under review. I have mentored over 20 undergraduate students, three of whom have since joined the Bioinformatics Interdepartmental PhD Program at UCLA. A complete list of my undergraduate mentees is available on my academic website: <http://www.sergheimangul.com/undegrads-in-genomics/>. My experience as a teacher in traditional classroom settings and as the mentor has been productive and gratifying; I look forward to expanding my teaching career through developing new computational courses to teach both computer science and biomedical researchers.