

## Involving undergraduates in genomics research to narrow the education–research gap

Sergei Mangul, Lana S Martin & Eleazar Eskin

**Engaging undergraduates in computational tasks can improve genomic research laboratory productivity, benefiting both students and senior laboratory members.**

The benefits of undergraduate involvement in scientific research have been well documented. Faculty supervision of undergraduate research projects provides for college students an invaluable, engaging, and challenging experience, unique to the research setting<sup>1</sup>. Undergraduates who engaged in research had higher grade point averages and higher rates of acceptance to graduate programs<sup>2</sup>. Students who completed an undergraduate research experience (URE) report an increased interest in pursuing STEM careers, understanding of how scientists work on problems, and ability to critically apply skills and interpret data<sup>3</sup>. The benefits of UREs are especially apparent for women and students of underrepresented minority (URM) groups. Underrepresented undergraduates who gain experience in STEM research are more likely to graduate from science and engineering majors, develop a more cohesive identity as scientists, and improve communication and critical thinking skills<sup>4</sup>.

### Challenges and opportunities of involving undergraduates

While the benefits of UREs for undergraduates are recognized, the advantages of UREs for graduate students, postdoctoral scholars, and faculty are not as clearly outlined (Fig. 1). A recent study identified a prevailing negative

attitude among faculty in STEM fields toward UREs; insufficient training, time, and incentive are among the barriers most commonly cited by faculty in regards to implementing or expanding research opportunities for undergraduates<sup>5</sup>. However, other studies have shown that involving undergraduates in laboratory research as an active learning component can actually strengthen research outcomes for senior lab members<sup>6,7</sup>.

In our computational genomics laboratory at the University of California, Los Angeles (UCLA), we have found that incorporating UREs benefits graduate students, postdoctoral scholars, and faculty. We 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 data sets 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<sup>8</sup>.

Senior researchers in the laboratory can directly benefit from undergraduate involvement in their projects. First, there is a growing need for high-throughput analyses in the field of genomics. Second, analysis of today's large data sets is time-consuming. Third, many tools emerge on a frequent basis that need to be installed, run, and tested. Undergraduates can acquire valuable training while enabling research laboratories to increase scalability of efforts to address these challenges. However, prevailing negative attitudes may deter

genomics faculty from incorporating UREs in laboratory research. Overcoming the research-and-teaching barrier requires a structured approach that recognizes the specific contributions that undergraduates can offer to genomics research.

### Framework for involving undergraduates in genomics research

Laboratories that perform interdisciplinary computational research are well-positioned to leverage the efforts of undergraduates while offering the student a valuable learning experience. Unlike wet labs, where training can be a slow process with numerous safety precautions, undergraduates in computational labs can be quickly and safely trained to produce publication-quality results. In computational genomics research, undergraduate trainees who master a particular skill can contribute sufficient work to gain authorship on a peer-reviewed paper.

We offer several tips for engaging undergraduates in genomics research while simultaneously improving laboratory productivity. First, identify 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-software tools is often an extremely complicated and time-consuming process—specifically when the software tool lacks detailed documentation.

Many research projects also require running a well-established computational pipeline for a large number of samples. While the computational capacities of the high-performance clusters allow the user to process a large amount of data, adjusting and supervising the pipeline might take a significant amount of the researcher's time. In many cases, these tasks

---

*Sergei Mangul, Lana S. Martin and Eleazar Eskin are in the Department of Computer Science, University of California, Los Angeles, Los Angeles, California, USA. Sergei Mangul is also at the Institute for Quantitative and Computational Biosciences, University of California, Los Angeles, Los Angeles, California, USA. Eleazar Eskin is also in the Department of Human Genetics, University of California, Los Angeles, Los Angeles, California, USA.*  
e-mail: [smangul@ucla.edu](mailto:smangul@ucla.edu)

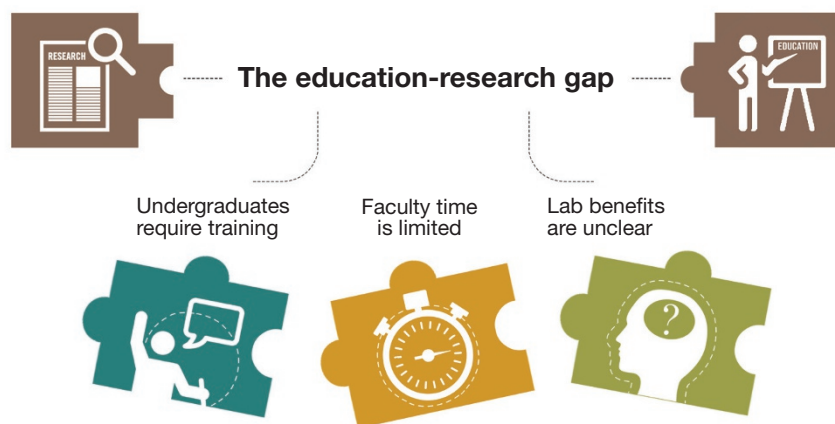


Figure 1 The education–research gap in universities.

are an ideal way for undergraduates to acquire hands-on skills in computer programming and process information about the applied subject, for example, biology, biochemistry, or genetics.

Second, encourage students to ‘outsource’ foundational education needs with 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 geared for first-time users. Lay-friendly review articles on otherwise complicated topics can help the undergraduate gain a basic understanding of the field without enrolling in a time-intensive course. Ideally, the student should be able to understand the ecological, biological, or medical problem that is the focus of the research without completing coursework on the subject.

Third, genomics research laboratories can take advantage of department- and campus-wide undergraduate research and training initiatives. For example, most undergraduates in our lab have benefited from two training pro-

grams offered by UCLA. The Bioinformatics Minor program allows students to develop an integrated understanding of contemporary genomic-scale research. Through a comprehensive inventory of courses, the minor in Bioinformatics provides a solid foundation in, and familiarity with, active research problems at the interface of computer science, biology, and mathematics. The Bruins-In-Genomics (B.I.G.) Summer is an intensive, practical experience in genomics and bioinformatics for students who are interested in integrating quantitative and biological knowledge, and pursuing graduate degrees in the biological, biomedical, or health sciences.

**Outcomes of undergraduate involvement**

We observed several substantial benefits that UREs bring to our computational genomics laboratory and our broader fields of study (Fig. 2). Delegating ‘low-level’ tasks to undergraduate students can unburden graduate students and postdoctoral scholars, freeing their time for ‘high-level’ work. This can allow the

researcher to simultaneously handle a larger number of projects. Encouraging undergraduates to seek extracurricular support for acquiring foundational skills both advances the student’s comprehension and technical skills while freeing time for faculty, graduate students, and postdocs—who might otherwise need to invest time for orientation of students. Finally, training undergraduates to perform research alongside postgraduate members of the lab can increase the pool of well-trained, high-performing scholars in the field.

As previously described in numerous studies, undergraduates gain substantial rewards from engaging in laboratory research as active learning. In our research laboratory, undergraduates gain co-authorship on papers for which they performed substantive low-level tasks and many students stay involved in projects well after the internship has ended<sup>9</sup>. Many students have gained admittance to competitive graduate programs in bioinformatics at UCLA and other universities. Our students come from fields that do not traditionally support authorship-generating UREs or provide computational training, such as major disciplines in the general life and medical sciences. After completion of the internship, these students are able to successfully perform computational tasks that prepare them for employment and competitive PhD programs.

**Conclusions, recommendations, and resources**

Based on our positive experiences mentoring undergraduate students in the laboratory, we are convinced that computational genomics research teams are ideally positioned to narrow the education–research gap. Given the simplicity and potential benefits reaped by senior researchers, our proposed strategy can be easily reproduced at other institutions, is pedagogically flexible, and is scalable from smaller to larger laboratory sizes. Our educational model is ideal for interdisciplinary, computational research units that employ computational approaches to analyze large-scale genomics data sets. While on-campus resources such as coding workshops and summer programs can help a laboratory scale up the number of undergraduates involved, research groups on smaller campuses with more limited resources can use online workshops to efficiently expand the pool of candidate URE participants<sup>10</sup>. We expect that an increase in hands-on training and research experiences in computational genomics for undergraduates from non-computational backgrounds will support greater integration of students from diverse

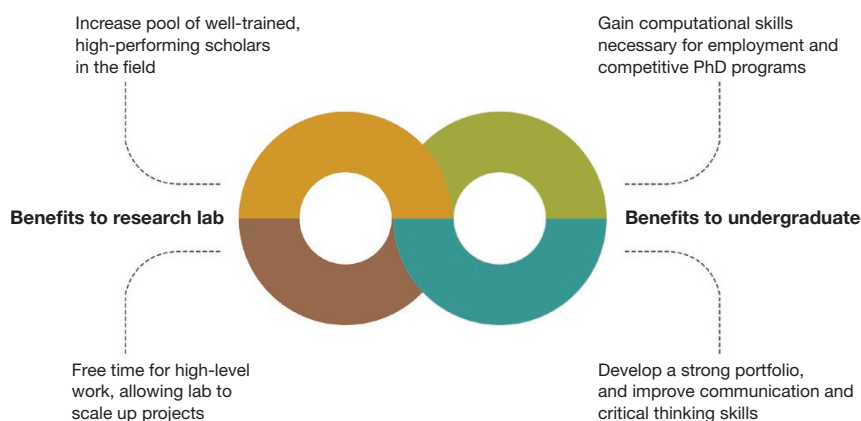


Figure 2 Benefits of UREs to the research laboratory and undergraduates.

backgrounds in science, technology, engineering, and mathematics careers.

#### ACKNOWLEDGMENTS

The authors thank the Bruins-In-Genomics (B.I.G.) Summer Undergraduate Research Program at the University of California, Los Angeles (UCLA) for supporting undergraduate students, and the Division of Undergraduate Education at UCLA for offering courses that allow undergraduates to work on independent research projects. Lastly, the authors dedicate this body of work to all undergraduate students who participated in

bioinformatics research in ZarLab (<http://zarlab.cs.ucla.edu>).

#### AUTHOR CONTRIBUTIONS

S.M. and L.S.M. contributed equally to the work in this manuscript.

#### COMPETING INTERESTS

The authors declare no competing interests.

1. McGee, R. & Keller, J.L. *CBE Life Sci. Educ.* **6**, 316–331 (2007).
2. Bauer, K.W. & Bennett, J.S. *J. High. Educ. (Columb. Ohio)* **74**, 210–230 (2003).
3. Hathaway, R.S. *et al. J. Coll. Student Dev.* **43**, 614–663 (2002).
4. Hernandez, P.R. *et al. J. Educ. Psychol.* **105**, 89 (2013).
5. Spell, R.M., *et al. CBE Life Sci. Educ.* **13**, 102–110 (2014).
6. Thomas, C. *Chronicle of Higher Education: the Conversation* <http://www.chronicle.com/blogs/conversation/2015/03/09/dont-divide-teaching-and-research/> (2017).
7. Druckman, J.N. *PS Polit. Sci. Polit.* **48**, 35–38 (2015).
8. Kerfeld, C.A. & Simons, R.W. *PLoS Biol.* **5**, e141 (2007).
9. Mangul, S. & Martin, L. *GitHub* <https://smangul1.github.io/undergraduates-in-genomics/> (2017).
10. Mangul, S. *et al. Trends Biotechnol.* **35**, 901–903 (2017).