

Work in Progress: Development of a Training Program to Prepare Students for an Immersive Bioinformatics Summer Research Experience

Prof. Mark A. Chapman, University of San Diego

Mark Chapman is an assistant professor at the University of San Diego in the Department of Integrated Engineering. His interests lie in the fields of skeletal muscle mechanics, muscle disease, exercise physiology, international education and engineering education. He earned his MS and PhD in bioengineering from the University of California, San Diego and a B.S. in biomedical engineering from the University of Minnesota.

WIP: Development of a training program for undergraduate students participating in an immersive bioinformatics summer research internship

The publication of the first human genome in 2001 transformed biomedical research [1,2]. Since then, an explosion of new sequencing technologies has required engineers and computer scientists to invent computational methods to analyze and interpret the ever-growing data. Now, large-scale biological data encompasses many types of ‘omics’ datasets, including genomes, transcriptomes, proteomes and metabolomes, and each of these new datasets has created a new set of analytical challenges [3–6]. To meet this need, the field of bioinformatics has expanded significantly, but there is still a large need for engineers and scientists to work in this inherently interdisciplinary field [7]. Properly trained bioinformaticians have expertise in computer science/engineering and understand the biological and medical context underlying their work [8]. Therefore, the development of robust bioinformatics training programs is critical to educate the next generation of bioinformaticians. Although undergraduate degree programs in bioinformatics exist, providing students with hands-on bioinformatics skills through immersive research experiences is necessary to prepare students for graduate work. Thus, this work describes the recently funded NSF – International Research Experience for Students (IRES) site: US-Sweden Clinical Bioinformatics Research Training Program targeted at training students from diverse educational backgrounds to prepare them for authentic bioinformatics research experiences. Given the inherent interdisciplinary nature of bioinformatics, it is extremely difficult to design a training program that prepares students from different backgrounds (computer science, bioengineering, computational biology, biology) to be successful in a bioinformatics research group. Therefore, this ‘Work-in-Progress’ describes the pre-departure training program developed for this IRES site and the initial lessons learned.

Program overview and virtual training series

During each of the three years of the IRES program, six rising juniors/seniors will be sent to Stockholm, Sweden for 10 weeks to conduct hands-on bioinformatics research at The Science for Life Laboratory (SciLifeLab). Criteria for inclusion in the program include: successful completion of a bioinformatics-related course, enrollment in a primarily undergraduate institution (PUI) in Southern California, and interest in pursuing graduate studies in bioinformatics. To ensure a diverse cohort of students, women, black, latinx, Native American, and LGBTQ+ students are especially encouraged to apply.

The facilities, research environment and research techniques at the foreign research site are unique in the world. SciLifeLab is home to over 1,500 researchers across more than 150 groups that collaborate to advance molecular biosciences. Specifically, SciLifeLab has positioned itself as a leader in next- generation sequencing techniques and bioinformatics. As such, the work performed by the undergraduate participants will be at the forefront of ‘omics’ research. Many of the -omics technologies and clinical data are newly developed and are not widely available outside of Sweden. Therefore, the students in this program will be exposed to scientific techniques not available at universities in the United States. Data and experiences from this program are applicable to many summer programs that are aiming to encourage and prepare undergraduate students from PUIs for enrollment in graduate programs as well as develop a diverse cohort of globally engaged scientists/engineers that seek career opportunities and collaborators throughout the world.

Prior to the start of the program, accepted students will be required to attend ten 90-minute weekly training sessions over Zoom. These training sessions will focus on material that is

suitable for virtual delivery. As such, these sessions will consist of lectures and activities centered around the culture of the host country (inside and outside the lab), genetics, genomics, and computer programming. In addition, this training program will include a journal club in which each student will present a paper from the lab they will be joining at SciLifeLab. Students will lead a discussion of the assigned article. This activity will help students understand the project they will work on over the summer and allow them to hit the ground running upon arrival to the host labs. Examples of planned activities for this virtual training series are outlined in Table 1.

Table 1: 10-week virtual training series – weekly 90-minute Zoom sessions

Week	Topic	Homework
1	Welcome & Introduction to Sweden and Swedish culture (inside and outside of the lab)	N/A
2	Genetics - DNA & RNA structure and function	Reading
3	Genetics - Gene expression	Reading
4	Genomics - next generation sequencing technologies	Reading
5	Genomics - next generation sequencing data	Reading
6	Sequencing data and the Unix Command Line	Coding Exercises
7	Data analysis in R	Coding Exercises
8	Journal Club	Article Reading
9	Journal Club	Article Reading
10	Journal Club	Article Reading

Pre-departure symposium

Following the 10-week training program, students will meet at the University of San Diego (USD) for a 4-day pre-departure symposium designed to prepare the IRES participants for their research projects. The topics covered during this symposium will begin with general principles that are applicable to all IRES participants and conclude with individualized coding sessions that are specific to each student's project (Table 2). The first day will focus on welcoming the students, discussing career paths in bioinformatics and include a special session from the on-campus international center. Since a main goal of this program is to encourage matriculation onto graduate school, a large portion of the first day will be dedicated to discussing graduate school admissions requirements and graduate school fellowships. In addition, the students will hear from a professional development panel about the possible career trajectories for bioengineers/bioinformaticians. Finally, the International Center at USD will host an informational session to prepare students for their time abroad. This session will include information on healthcare, stress management, culture shock, and safety tips while living abroad.

Table 2: Example Pre-departure Bioinformatics Symposium Schedule

Day	Topics
1	Greetings, USD International Center, career panel, graduate school and fellowships
2	Accessing/Importing sequencing data & data quality control
3	Read mapping of sequencing data & differential gene expression
4	Individualized bioinformatics training

The remaining three days of the pre-departure symposium will focus on bioinformatics skills necessary for student success during their research projects. Since all students will be working with sequencing datasets, common topics will include: accessing/importing sequencing data, data quality control, read mapping, differential gene expression and gene ontology. On the final day of the symposium, students will work on bioinformatic techniques that are specific to their summer projects. Specifically, they will use this time to familiarize themselves with the bioinformatic packages currently in use at their assigned laboratories. These individualized training sessions have been designed together with the group leaders at the foreign host lab to ensure students are ready to provide meaningful contributions to the host lab. Following this

symposium, the IRES students will travel together to Stockholm to perform research at the host laboratories. The pre-departure symposium will also include social activities so that the students bond prior to their departure to Sweden. Coffee breaks, social lunches and a group dinner will help the students develop a good rapport before spending the summer together abroad.

Research development workshops:

After the pre-departure symposium, the students will travel together to Stockholm to begin their research projects at SciLifeLab. The author will travel with the students and be onsite during the first 3 weeks of the program. During these first three weeks at the host institution, weekly 2-hour research development workshops are to be held (Table 3). These workshops are designed to provide students with information on how to be an effective and ethical scientist. Since many of the student participants will be new to scientific research, workshop topics include: the scientific method, hypothesis formulation, searching the scientific literature, managing references, writing scientific reports, experimental design, laboratory documentation (lab notebooks and data management), research ethics, presentation skills, and authorship guidelines.

Table 3: On-site Research Development Workshops – Weekly 2-hour sessions

Week	Topics
1	Searching the scientific literature, scientific method, hypothesis formulation
2	Laboratory documentation, data management, research ethics
3	Writing scientific reports, presentation skills

Additionally, during the first weeks of the program, the author will meet with each student, their direct supervisor (graduate student/postdoc), and their research group leader (as needed) to discuss the students' progress and address any concerns that may have arisen. The author will not remain onsite during the entire 10-week program, so the initial on-site meetings will be replaced with one-on-one video conference mentorship meetings that occur weekly for the remainder of the summer. Prior to each of these meetings, the student will send an email detailing their accomplishments for that week and experimental plans for the upcoming week. Additionally, the student will be asked to share a personal reflection about their time abroad in this weekly email update.

Additional personalized training

It is recognized that even with this training program, some students may still struggle with aspects of their research projects. As such, additional one-on-one training will take place on an individual basis as needed. This is to take place over Zoom in the weeks leading up to the program as well as throughout the summer.

Continual improvement

In order to ensure continual improvement, the program's external evaluation will include an assessment of the above training strategies. Specifically, feedback will be collected from the students to assess how prepared they felt from a technical standpoint going into their labs. This information will be used to determine if there are any aspects of the technical training sessions that need to be altered for future cohorts. Additionally, students' perceptions of their personal preparedness (i.e. issues related to culture shock and stress management) will also be assessed to determine if changes need to be made for future cohorts. Finally, feedback will be gathered to assess the quality of the program's professional development workshops. Feedback from host laboratories will also be sought to evaluate perceived technical and personal student preparedness. These data are to be presented at the ASEE conference and should be of particular interest to directors of other undergraduate summer research programs.

References

1. Lander ES, Linton L, Birren B, Nusbaum C, Zody M, Baldwin J, et al. Initial sequencing and analysis of the human genome. *Nature*. 2001;409:860–921.
2. Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG, et al. The sequence of the human genome. *Science* (80-). 2001;291:1304–51.
3. Kircher M, Kelso J. High-throughput DNA sequencing - Concepts and limitations. *BioEssays*. 2010;32(6):524–36.
4. Gross M. Riding the wave of biological data. *Curr Biol*. 2011;21(6):R204–6.
5. Lindblom A, Robinson PN. Bioinformatics for Human Genetics: Promises and Challenges. *Hum Mutat*. 2011;32(5):495–500.
6. Tenenbaum JD. Translational Bioinformatics: Past, Present, and Future. *Genomics, Proteomics Bioinforma*. 2016;14(1):31–41.
7. Gómez-López G, Dopazo J, Cigudosa JC, Valencia A, Al-Shahrour F. Precision medicine needs pioneering clinical bioinformaticians. *Brief Bioinform*. 2017;1–15.
8. Schneider MV, Watson J, Attwood T, Rother K, Budd A, McDowall J, et al. Bioinformatics training: A review of challenges, actions and support requirements. *Brief Bioinform*. 2010;11(6):544–51.