RESEARCH ARTICLE

Bioinformatics Bootcamp: A model for training clinical researchers

Authors

Debra Murray, Ph.D.[1], Jennifer Drummond, M.S.[2], Deborah Ritter, Ph.D.[1], Kirt Martin, Ph.D.[1], Lori Banks, Ph.D. [3], Dewayne Whittington[4]

Affiliations

[1] Baylor College of Medicine

[2] Rice University

[3] Bates College

[4] Strategic Evaluations, Inc.

Correspondence

Debra Murray

Email: ddm@bcm.edu

Abstract

Within the last few decades, there has been a steady decline in physicians conducting research. The U.S. Biomedical workforce is aging which affects improving health care in this country. As they age, we need innovative ways to increase the physician-scientist pool. Along with shortages of under-represented minorities in this field, there is a scarcity of people able to analyze biological data. In an effort to increase the number of under-represented minorities conducting bioinformatics/genomics research, the Human Genome Sequencing Center-Genetics/Genomics Research Education Training (G/GREAT) program created a bioinformatics mini course. The Bioinformatics Boot camp course offers an opportunity to learn a skill that will increase summer interns' self-confidence, interest, and proficiency in seeking future computational research opportunities. This study will determine the best training approach to accomplish this goal for the novice (non-experienced) student. We surveyed course instructors to understand their bioinformatics teaching method and G/GREAT interns for their perspectives concerning the course. The results provided value information that guided curriculum design for all future G/GREAT training courses. These outcomes suggest that similar courses aimed at clinicians interested in research could increase physician scientists to begin replacing those about to retire.

Terms: Interns=students=participants, programming=coding, non-experienced=novice.

1.0 Introduction

Within the last few decades, there has been a steady decline in physicians conducting research[1]. Physician-scientists are well prepared to detect new threats to human health; develop potential new therapies, treatments, or means of prevention; communicate knowledgeably across disciplines and to lead scientific teams or organizations; and guide important policy decisions[2]. However, the U.S. biomedical workforce is aging, which affects improving health care in this country. The current physician-scientist workforce is being sustained by an aging group[3,4]. The National Institutes of Health (NIH) reports that the number of younger physician-scientists applying for research support is decreasing and that the average age of these investigators, including first-time applicants, is increasing. The Physician-Scientist Workforce working group report data was generated from NIH and American Medical Association (AMA) surveys. The NIH Advisory Committee to the Director (ACD) analyzed data on M.D./Ph.Ds., M.Ds., nurses, and other researchers with clinical training collected from an AMA survey finding that the number of physicians conducting research has declined 5.5% since 2003 to about 13,700 in 2012[1]. The lack of diversity of the physician-scientist workforce, along with this shortage, is also a source of very serious concern to the NIH[1]. The working group made recommendations similar to those of the 2012 Biomedical Workforce Working Group of the ACD, like enrich training programs and give more weight to proposals from young researchers, as well as new ones by creating a category for physician-scientists within the K99/R00, awards[3,5].

The field of Bioinformatics grew rapidly out of the Human Genome Project. It is a subdiscipline of biology and computer science concerned with the acquisition, storage, analysis, and dissemination of biological data, most often DNA and amino acid sequences[6]. Bioinformatics uses computer programs for a variety of applications (gene identification and sequence analyses, prediction of protein structures, genome annotation, comparative genomics, health and drug discovery) to manage volumes of biological information. In recent years, for example, successful translations of genetic and genomic research have seen significant investments in rapid advances in large-scale genomics and related technologies into cardiovascular disorders, pulmonary diseases, and blood disorders, with discoveries of genes underlying chronic hypertension, congenital heart disease, aortic aneurysm, idiopathic pulmonary fibrosis, and modifiers of sickle cell disease, to name but a few. These discoveries have provided critical insights to the underlying biology of these disorders and thus stand as a foundation for future therapeutic initiatives. Encouraging more clinicians to take advantage of this research field could provide a solution to the U.S. health care crisis.

The authors suggest that bioinformatics development among clinicians as a solution to both the decline in medical researchers and the need for more skilled physicians to analyze the large amounts of data generated by genomic studies. This focus would increase knowledge among physicians concerning genetics and genomics needed in the clinic. Because of the advances genetics and genomics have made, the NIH, CDC, and the Health Services and Resources Administration convened a workshop to identify practical strategies to educate primary care physicians involved in maternal and child health[7]. The advancements in genetic screenings will, in large part, require primary care physicians to increase understanding of genetics and genomics to help the patient interpret findings that may affect their care throughout their lifetime.

The authors developed a mini course to address under-represented minority students' hesitations concerning training in bioinformatics. There are many online coding sites that teach programming languages useful in bioinformatics research. For example, Codeacademy.com[8] offers online interactive courses to learn C++, JavaScript, and Python, to name a few; and Coursera.org[9] offers free-based programming courses taught by instructors from institutions like Harvard and MIT, that are pre-recorded video lectures, auto-graded with peer-reviewed assignments. Although these and more courses exist, few offer the pre-introductory information that the non-experienced coder will need to complete these courses successfully. The mini course, Bioinformatics BootCamp, provides a useful tool to increase UR minorities' self-confidence to train in this growing field. This study describes the development of the mini course as an example to increase more researchers with bioinformatics skills. We offer this approach as a way to increase more clinicians with bioinformatics skills to pursue research careers.

1.1 Educational Environment

1.1.1 Institution

For 75 years, Baylor College of Medicine (BCM) has provided medical and research training in the state of Texas. It is the highest ranked medical school in the state and the Southwest region. BCM is among the top 15% medical schools in the nation. It is ranked 22nd for research-intensive medical schools and 4th in primary care in the 2020-2021 U.S. News & World Report annual list of top graduate schools. Nationally, the Graduate School of Biomedical Sciences (GSBS) ranks 26th in the biological sciences, and the Baylor Physician Assistant program ranks 3rd. The nurse anesthetist doctoral program ranks 2nd among all masters/doctorate nurse anesthesia programs in the country. The Department of Molecular and Human Genetics (MHG) and the Human Genome Sequencing Center (HGSC) are world-leaders in genetics and genomics and has played an active role in advancing genomics. Established in 1985, the MHG is ranked #1 among Medical Schools in NIH-funding for genetics. The HGSC is a world-leader in genomic technology development and genomic sequencing, having sequenced more than 20,000 human whole-exomes and whole-genomes. Because of the innovation required to solve difficult sequencing challenges, the HGSC has developed significant software, technologies, and tools for data analysis that has ultimately led to establishing a CAP/CLIA-certified clinical laboratory to enable next-generation sequencing efforts for research studies requiring clinically-validated methods and return of clinical results.

1.1.2 HGSC-G/GREAT Program

The HGSC-Genetics/Genomics Research Education and Training (G/GREAT) program is an undergraduate summer research internship. The HGSC-G/GREAT program is a part of the HGSC-Diversity Initiative to increase STEM under-represented groups in genetics/genomics with Ph.Ds. and M.D./Ph.Ds. Rivera and Murray (2014) reported on participants in this program described that expressed that the program assisted them in gaining concrete experiences in science research, learning about options in science, and learning how to continue in a science academic path[10]. The summer research program provides mentored research training in bioinformatics/genetics/

genomics, GRE preparation course, graduate school career preparation, program courses, professional development seminars, responsible conduct in research training, biomedical research seminars, and mentoring. The undergraduate research program has a strong record of training and preparing undergraduates for graduate programs. The most recent R25 program data (2011-2016) show that 48% entered Ph.D. programs, while 11% enrolled in Master's Programs. Since the inception of this program in 2003, 100% of the undergraduates obtain bachelor's degrees in STEM and also entered biomedical related careers. HGSC-G/GREAT program demographic data is as follows: 66% African American; 30% Hispanic American; and 4% Caucasian; with 55% economically disadvantaged, and 0% persons with disabilities; 23% males and 77% females; and a mean GPA of 3.3.

2.0 Methods

2.1 Research Questions

What bioinformatics teaching method is the best approach to interest the novice student?

What teaching methods were most promising in increasing the interest, self-confidence, and introductory bioinformatics proficiency levels among novice students?

2.2 Study Population

The HGSC-G/GREAT program accepts undergraduate students from groups under-represented in genomics (African Americans, Latino Americans, Native Americans, Pacific Islanders, economically disadvantaged, and Persons with Disabilities). Students were expected to have at least a 3.0 cumulative grade point average (GPA) and declared a STEM major in biology, chemistry, computer science, math, mechanical or electrical engineering. Program participants were assigned research projects in bioinformatics, molecular genetics, genomics, and engineering. Table 1 lists the demographics of the study cohorts. The STEM majors varied among the groups. In 2013 the college majors of this cohort included biology, biology/premed, biology/chemistry/Spanish, clinical laboratory science, and mathematical biology/chemistry. The average GPA was 3.6. The present educational outcomes of this cohort are 1 J.D., 1 M.D., 1 M.S., and 2 Ph.Ds. The first G/GREAT alumni awarded a Ph.D. in Bioinformatics came from this group. The 2014 cohort had a 3.2 average GPA with majors in biochemistry, biochemistry/philosophy, biology/chemistry, biology/education, computer science, mathematics, mechanical engineering. Currently, this group has one 1st year medical student, 3 in Ph.D. programs, 1 engineer, 1 M.S., and 1 programmer. The 2015 group had the following majors: biochemistry/ chemistry, biology, biochemistry, bio-chemistry/applied mathematics, cell and molecular/chemistry, and biology/secondary education. The average cohort GPA was 3.4. The 2015 alumni have achieved the following 1 M.S. and 5 in Ph.D. programs.

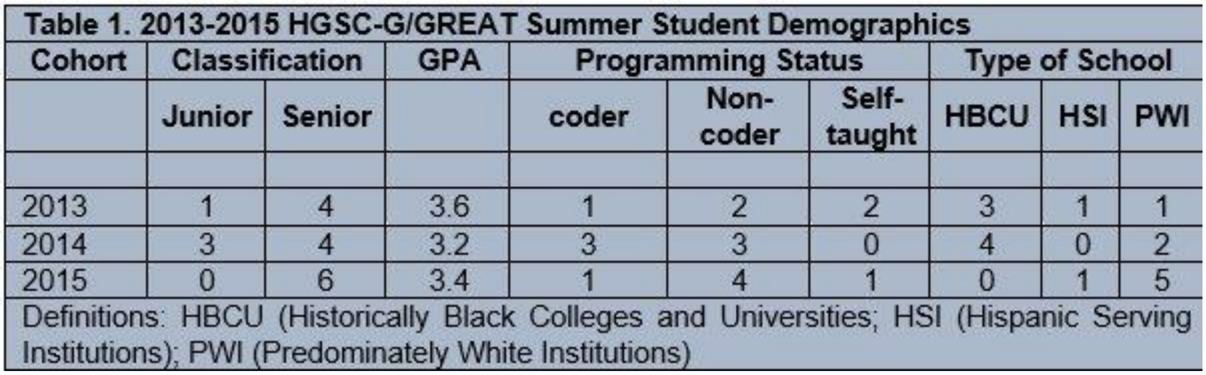


Table 1. 2013-2015 HGSC-G/GREAT Summer Student Demographics

Cohort	Classification		GPA	Programming Status			Type of School		
	Junior	Senior		coder	Non-coder	Self-taught	HBCU	HSI	PWI
2013	1	4	3.6	1	2	2	3	1	1
2014	3	4	3.2	3	3	0	4	0	2
2015	0	6	3.4	1	4	1	0	1	5

Definitions: HBCU (Historically Black Colleges and Universities; HSI (Hispanic Serving Institutions); PWI (Predominately White Institutions)

2.3 Survey Design and Instrumentation

This work was carried out under the approval of the Institutional Review Board (IRB) for Human Subject Research for Baylor College of Medicine and Affiliated Hospitals (BCM IRB) approved H-27027. All participants provided consent for data collection used in this study. G/GREAT participants completed online surveys to assess perspectives concerning the entire program with an evaluator that also conducts in-person focus group interviews. The survey included a section with a set of questions concerning the bioinformatics BootCamp course. Survey response rate was 100%. A short pre/post survey was given before and after the mini course to determine participants' familiarity with bioinformatics concepts. The course survey was not given in 2014. The program evaluator analyzed the responses and submitted a report to the program leadership. The instructor survey was collaboratively created by program leadership and evaluator to determine the design elements that are most important for instructors to consider when building a pre-introductory bioinformatics course to train a novice (non-experienced coder) student. In addition, the survey aimed to document promising teaching methods for increasing the interest, self-confidence levels, and introductory-level bioinformatics proficiency among novice students. This survey was administered to the instructors involved in teaching the course during 2013, 2014, and 2015. There was a 100% response rate.

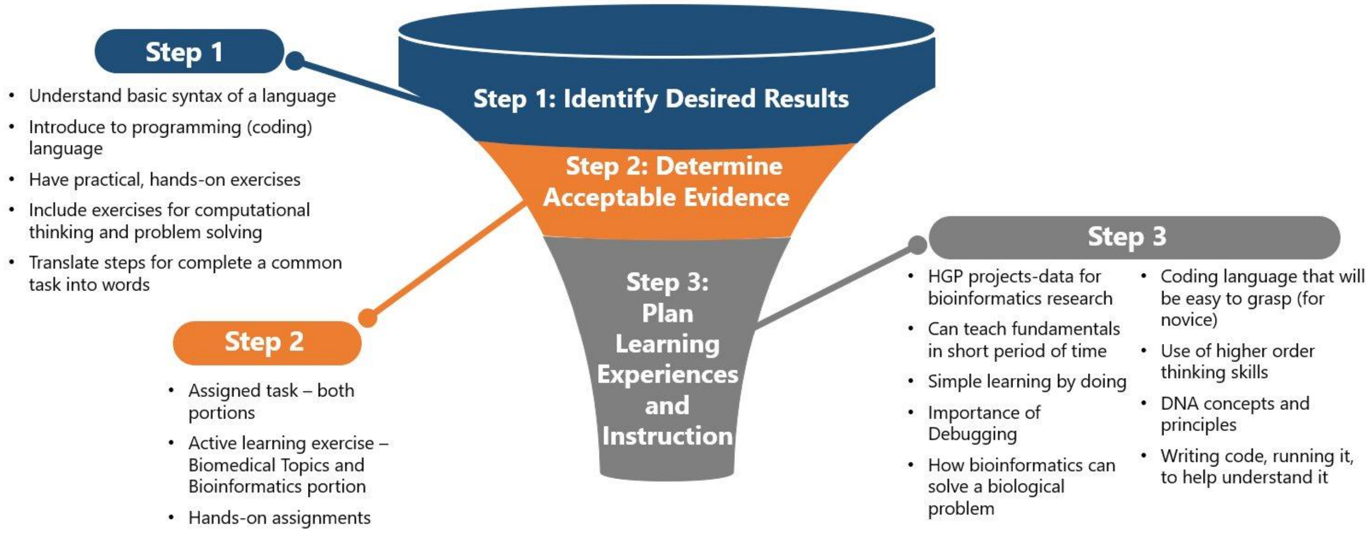
2.4 Data Analysis

Data were analyzed in collaboration with Strategic Evaluations, Inc. (Durham, NC). All descriptive data tables and statistical tests were performed using *SPSS v21* (IBM, New York, NY) to compute descriptive statistics, as well as test for correlations and statistical significance. Qualitative data submitted for open-ended questionnaire responses were analyzed via *Atlas.ti* (Berlin, Germany) to assign thematic codes to participants' narrative comments.

2.5 Backward Design of Course Curriculum

As a result of our experience in this area to increase the likelihood that UR students will use bioinformatics in their research and future studies, we developed a bioinformatics mini course to introduce this field. We used the *Backward Design* course model to develop a bioinformatics course for the G/GREAT interns. Wiggins and McTighe (2005) state curriculum should lay out the most effective ways of achieving specific results. Answering the "why?" and "so what?" questions that older students always ask (or want to) and doing so in concrete terms as the focus of curriculum planning is thus the essence of understanding by design[11]. The G/GREAT program used the outcome of ‘introducing bioinformatics to the interns to increase their interest and self-confidence’. Using the three states of Backward Design (1. Identify desired results, 2. Determine acceptable evidence and 3. Plan learning experiences and instruction), the leadership scrutinized this model to develop curriculum for a bioinformatics mini course (Figure 1). It was important for this course to include opportunities for computer science majors to understand biological concepts, life science majors training in bioinformatics concepts, and both majors to learn bioinformatics tools. It was also critical for the biomedical topics portion of the course to include active learning modules to help the students engage with the material.

Figure 1: HGSC-G/GREAT *Backward Design* Model. The Backward Design Model was used to create the curriculum for the Bioinformatics Bootcamp course.


Step 1
Understand basic syntax of a language
Introduce to programming (coding) language
Have practical, hands-on exercises
Include exercises for computational thinking and problem solving
Translate steps for complete a common task into words
Step 1: Identify Desired Results
Step 2: Determine Acceptable Evidence
Step 3: Plan Learning Experiences and Instruction
Step 2
Assigned task – both portions
Active learning exercise – Biomedical Topics and Bioinformatics portion
Hands-on assignments
Step 3
HGP projects-data for bioinformatics research
Can teach fundamentals in short period of time
Simple learning by doing
Importance of Debugging
How bioinformatics can solve a biological problem
Coding language that will be easy to grasp (for novice)
Use of higher order thinking skills
DNA concepts and principles
Writing code, running it, to help understand it

2.6 Course Description

The *Bioinformatics BootCamp* course was designed to make *bioinformatics concepts* familiar to students unacquainted with this field. This course aimed to introduce under-represented students to various aspects of the field (comparative genomics/coding) and provide them with a way to address the challenges. This mini course also expected to increase the UR students’ confidence in their abilities to code and reduce doubt regarding their skills to pursue career options in this field. The course was offered four days a week for two hours during the first week of the summer program. The students were not required to have prior computational knowledge to take the course. The course was divided into two components: one that focused on biomedical topics and another segment that introduced students to bioinformatics. The program director taught the biomedical topics component and addressed topics such as DNA chemistry and structure, understanding mutations, introduction to relevant projects like HAP MAP[12], 1000 Genomes[13], All of US[14]. HGSC bioinformaticians taught the bioinformatics sessions, which provided the history, concepts, and programming activities. Although the course was designed for the

G/GREAT program, it was not limited to this group. Summer students from the larger BCM internship program (especially interns assigned in the HGSC) and BCM graduate students were also able to take the course.

2.7 Course Teaching Team

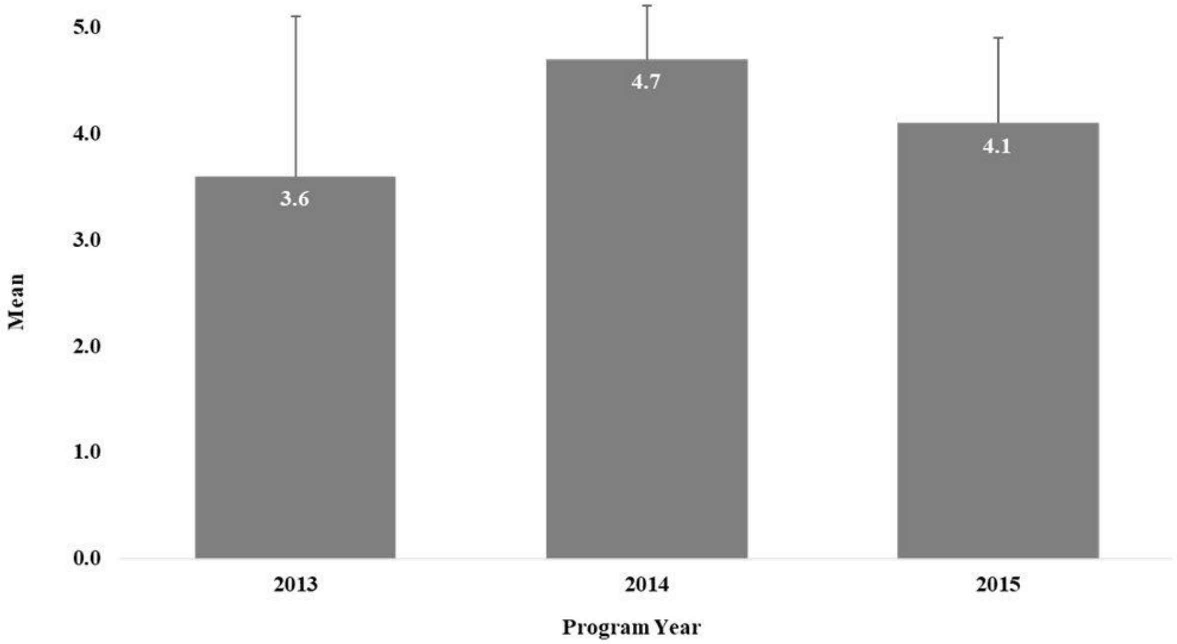
Because the make-up the G/GREAT interns varied from year to year, the course included a molecular biology instructional component for the undergraduates who were majoring in computer science and/or engineering. The program director taught this portion of the course. This instructor has a Ph.D. in Cell and Molecular Biology and spent ten years prior to designing the course in Genetics and Genomics education. With teaching experience at both majority and minority-serving institutions, this instructor has developed several courses for the program. The bioinformatics instructors all trained in the field of bioinformatics and had conducted research in this field. Each also had teaching experience. The first bioinformatics instructor (A) has a Ph.D. in Computational Biology and primary research areas were in bioinformatics. Instructor A also taught undergraduates at local colleges for several years. The second bioinformatics instructor (B) was an Institutional Research and Career Development Award (IRACDA) post-doctoral fellow at BCM with a Ph.D. in Biology and Bioinformatics. The IRACDA postdoctoral fellowship allows trainees time to develop teaching skills while maintaining a focus on their research. The last bioinformatics instructor (C) has a M.S. in Bioinformatics and was a bioinformatician in the HGSC. Instructor C trained newly hired programmers in the center to adjust to the genomic projects they were assigned. An additional instructor for this course series provided a session that discussed journal articles and utilized data generated by instructor C. This instructor teaches the molecular genetics review course for the G/GREAT program and has a Ph.D. in Microbiology and Biochemistry and was also an IRACDA fellow.

3.0 Results

In the early years of the HGSC-G/GREAT program, students placed in bioinformatics labs were unable to have a meaningful experience with bioinformatics and/or generate data. After three years, five students had been placed in bioinformatics projects, and only two could be counted as mildly successful. These two students persevered and wanted to conquer the research problem they were given and remained on the project to the end of the internship. These two students were males. The five students had the following make-up: 60% female, 40% male (2), and 40% African American and 60% Hispanic American. Because of this experience, we focused on providing bioinformatics training opportunities to eliminate students' fears of programming and increase their willingness to train in this research area.

This experience led the program leadership to consider ways to prepare G/GREAT interns that will ensure success in bioinformatics projects. We needed the curriculum to address three groups: (1) non-experienced (also called novice) students, (2) non-biology majors (computer science, physics, statistics), and the highly motivated students (self-taught coders). We used the *Understanding by Design* concept to develop the curriculum for this mini course[11]. This course would introduce key concepts that prepare students for research assignments that included bioinformatics. The curriculum included 50% DNA concepts/topics and 50% introductory bioinformatics. DNA concepts/ topics (DNA chemistry and structure, mutations, SNP marker) provided basic understanding for biology/chemistry majors as a refresher and served as an introduction for computer science/ engineer-ing/physics majors. The students learned about significant large-scale DNA projects that would help them in future bioinformatics tasks like HAPMAP[12], 1000 Genomes[13], All of Us[14]. Over the course of the first three years, the bioinformatics portion was taught by a different instructor. We surveyed the G/GREAT interns to document their perspectives concerning all of the program activities. During the first three years of the bootcamp course, the students indicated that this mini course was a benefit to their career development (Figure 2). The 2013 cohort trended toward neutral when rating the benefit of the bootcamp (Figure 2). But the following two years achieved more substantial mean ratings for the benefit of the course (Figure 2). For example, the 2014 G/GREAT cohort rated the bootcamp highly beneficial, with a mean rating of 4.7 on a 5-point Likert scale. Students in the 2015 group also submitted strong mean scores for the benefits of the course, with a mean rating of 4.1.

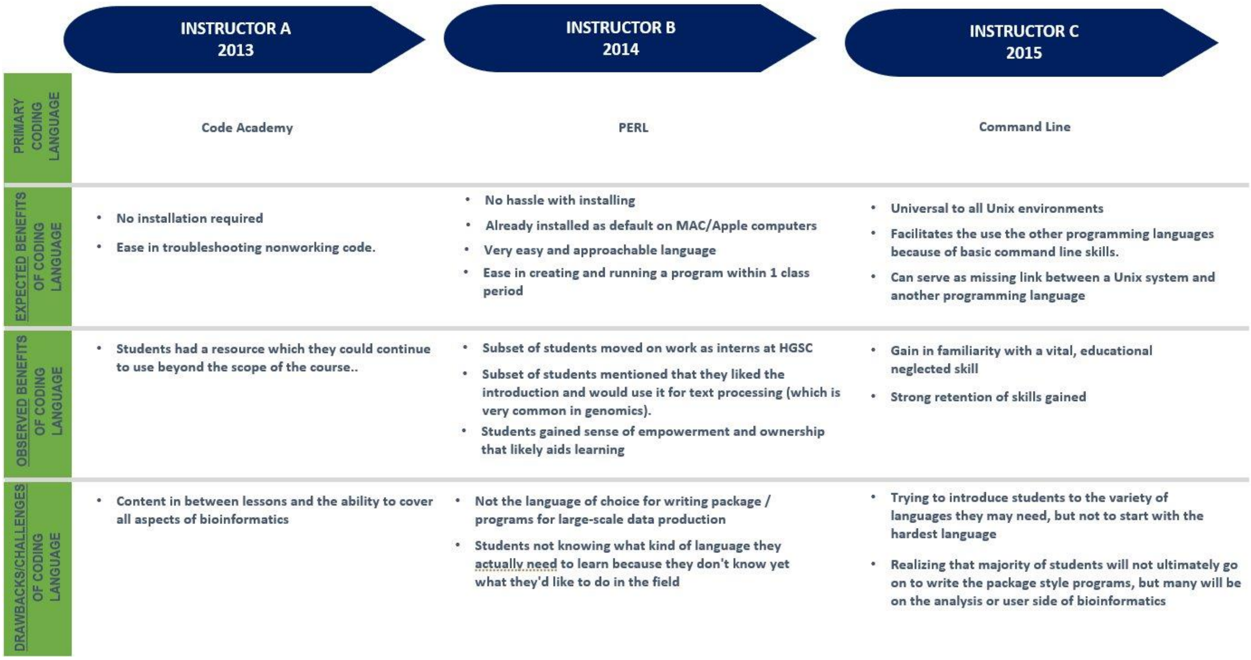
Figure 2: Students' Mean Ratings for Extent to Which Bioinformatics Bootcamp was Beneficial. Survey data were collected over a three-year period (2013-2015), with students rating the benefit of the bioinformatics bootcamp on a 5-point Likert scale. Bars in the figure represent the mean rating and standard deviation for each year.



*A higher mean indicates a higher level of students rating the bioinformatics bootcamp beneficial, as 1 equaled "Not Beneficial" and 5 equaled "Extremely Beneficial"

Since we developed this course to encourage the non-experienced students to increase self-confidence in taking on bioinformatics projects, we wanted to investigate the instructors' approach to the bioinformatics portion of the mini course. We surveyed the instructors for feedback, and their responses are found in Figures 3 and 4. Figure 3 reports each instructor's choice of programming language (also called 'coding') and the rationale for its use. The instructors used a variety of approaches to introduce various programming languages for the study of bioinformatics.

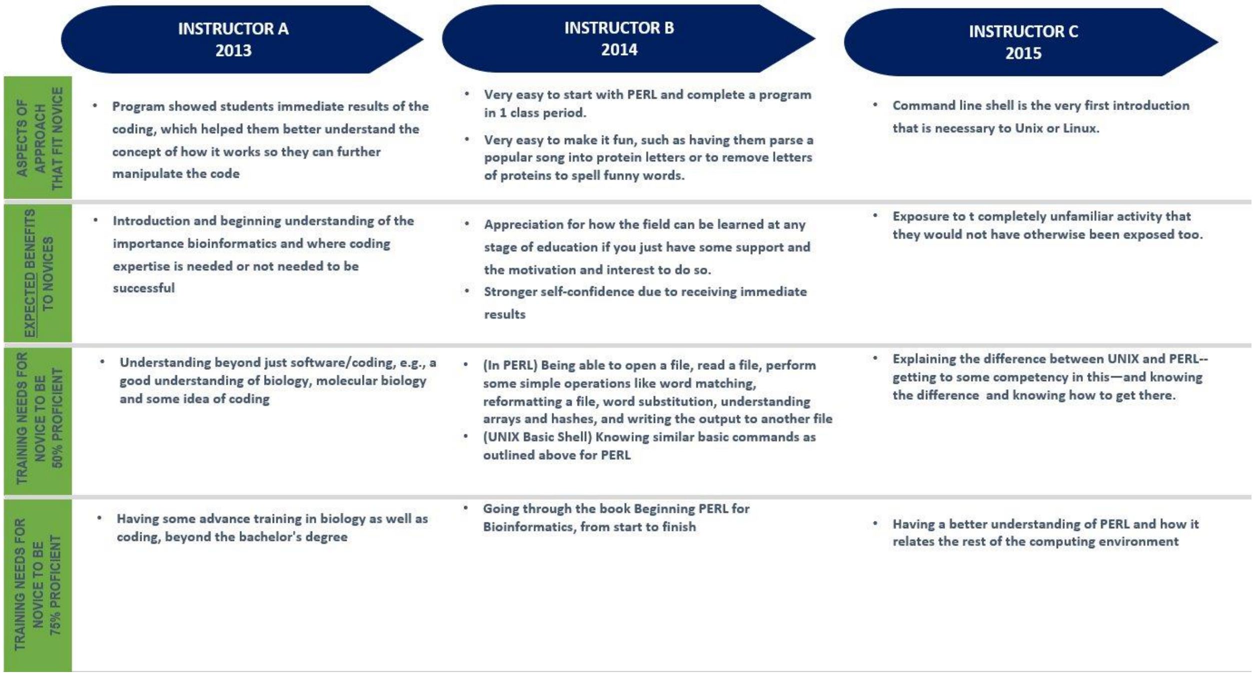
Figure 3: Instructors' Survey. Choice of programming language for Bioinformatics course.



	INSTRUCTOR A 2013	INSTRUCTOR B 2014	INSTRUCTOR C 2015
PRIMARY CODING LANGUAGE	Code Academy	PERL	Command Line
EXPECTED BENEFITS OF CODING LANGUAGE	• No installation required • Ease in troubleshooting nonworking code.	• No hassle with installing • Already installed as default on MAC/Apple computers • Very easy and approachable language • Ease in creating and running a program within 1 class period	• Universal to all Unix environments • Facilitates the use the other programming languages because of basic command line skills. • Can serve as missing link between a Unix system and another programming language
OBSERVED BENEFITS OF CODING LANGUAGE	• Students had a resource which they could continue to use beyond the scope of the course..	• Subset of students moved on work as interns at HGSC • Subset of students mentioned that they liked the introduction and would use it for text processing (which is very common in genomics). • Students gained sense of empowerment and ownership that likely aids learning	• Gain in familiarity with a vital, educational neglected skill • Strong retention of skills gained
DRAWBACKS/CHALLENGES OF CODING LANGUAGE	• Content in between lessons and the ability to cover all aspects of bioinformatics	• Not the language of choice for writing package / programs for large-scale data production • Students not knowing what kind of language they actually need to learn because they don't know yet what they'd like to do in the field	• Trying to introduce students to the variety of languages they may need, but not to start with the hardest language • Realizing that majority of students will not ultimately go on to write the package style programs, but many will be on the analysis or user side of bioinformatics

They determined that their approach would have the following benefit to the interns: Code Academy did not require installation of software and there was a way to troubleshoot nonworking code (Instructor A); PERL which is already installed on MAC/Apple computers, was an easy and approachable language, and students could create and run a PERL program within one class period (Instructor B); and the Unix command line was universal to all Unix environments and often the missing link between a Unix system and a programming language (Instructor C). The actual benefits of the programming choices to the class as witnessed by the instructors were: (1) students had a resource they could continue to use beyond the course, (2) a few of the students interned at HGSC had prior coding experience but not in PERL, and (3) students gained familiarity in a skill that was vital and typically neglected. All of the instructors agreed that the non-experienced student could learn from their respective programming approaches (Figure 4). Instructor A said it gave the students a beginning; an introduction taught them that this is an important aspect of it because you really don't have to code to do bioinformatics, it is a benefit. Instructor B said students especially appreciated the boot camp because it was an example of how this field can be learned at any stage of education if you just have some support and the motivation and interest to do so. Novice students especially liked quickly starting with immediate results (creating a program, running it or finding a database and getting data and doing some operation on it). Lastly, Instructor C thought the approach served the non-experienced students pretty well because it was a completely unfamiliar activity that they would not have otherwise been exposed too. The instructors believed that this immediate gratification helped build students' self-confidence.

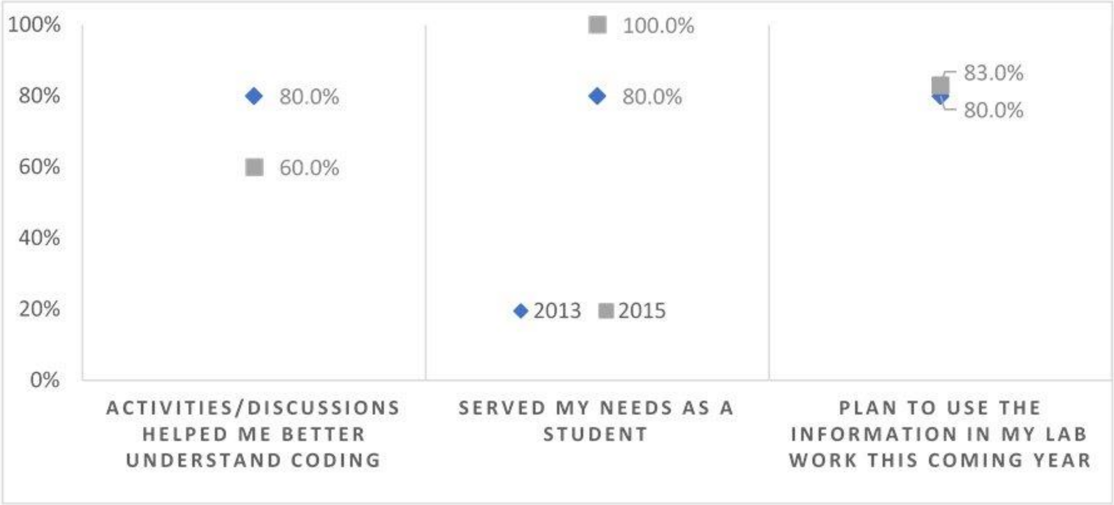
Figure 4: Instructors' Survey. Ability for novice to use chosen computer language.



	INSTRUCTOR A 2013	INSTRUCTOR B 2014	INSTRUCTOR C 2015
ASPECTS OF APPROACH THAT FIT NOVICE	• Program showed students immediate results of the coding, which helped them better understand the concept of how it works so they can further manipulate the code	• Very easy to start with PERL and complete a program in 1 class period. • Very easy to make it fun, such as having them parse a popular song into protein letters or to remove letters of proteins to spell funny words.	• Command line shell is the very first introduction that is necessary to Unix or Linux.
EXPECTED BENEFITS TO NOVICES	• Introduction and beginning understanding of the importance bioinformatics and where coding expertise is needed or not needed to be successful	• Appreciation for how the field can be learned at any stage of education if you just have some support and the motivation and interest to do so. • Stronger self-confidence due to receiving immediate results	• Exposure to t completely unfamiliar activity that they would not have otherwise been exposed too.
TRAINING NEEDS FOR NOVICE TO BE 50% PROFICIENT	• Understanding beyond just software/coding, e.g., a good understanding of biology, molecular biology and some idea of coding	• (In PERL) Being able to open a file, read a file, perform some simple operations like word matching, reformatting a file, word substitution, understanding arrays and hashes, and writing the output to another file • (UNIX Basic Shell) Knowing similar basic commands as outlined above for PERL	• Explaining the difference between UNIX and PERL--getting to some competency in this—and knowing the difference and knowing how to get there.
TRAINING NEEDS FOR NOVICE TO BE 75% PROFICIENT	• Having some advance training in biology as well as coding, beyond the bachelor's degree	• Going through the book Beginning PERL for Bioinformatics, from start to finish	• Having a better understanding of PERL and how it relates the rest of the computing environment

We surveyed the 2013, 2015 G/GREAT interns with specific questions regarding their experience in the mini course. We do not have student responses from the 2014 course to include for this study. Figure 5 reports data concerning the students' attitudes toward the mini course. We asked if the activities/ discussions during the short course were effective in helping to better understand coding. Eighty percent (80%) of the 2013 cohort agreed or strongly agreed that the course achieved this goal, while 60% of the 2015 group felt similarly. Both the 2013 (80%) and 2015 (100%) cohorts agreed or strongly agreed that the short course was interesting and conducted in a manner that served their needs. We also wanted to know if the students planned to use the information provided in the mini course in their labs that year. Both the 2013 and 2015 cohorts highly rated the utility of the content for future laboratory work with at least 80%,83% (respectively) of interns in each group agreeing that they would use the content learned during the bootcamp in the upcoming year. There was a small percentage (up to 17%) of the interns who were not sure if they would use this training.

Figure 5: Students' Level of Agreement with Statements Connected to Bioinformatics Bootcamp. Survey data were collected over a two-year period (2013 & 2015), with students rating their level of agreement with three statements related to the impact and usefulness of the bioinformatics bootcamp. Ratings were submitted along a 5-point Likert scale, where "1" equaled "Strongly Disagree" and "5" equaled "Strongly Agree". Chart displays the percent of students who agreed or strongly agreed with each statement.



*survey for 2014 was altered, and these items were not rated by students

4.0 Discussion

With the advent of Next-Gen (Generation) Sequencing Technologies, a paradigm shift has occurred in that the rate-limiting step for genomic research is no longer work associated with the wet bench, but rather the handling and analysis of large amounts of data. It is becoming more cost effective to generate genomic data, but we do not have the workforce to translate this information into knowledge. Chang (2015) reports that biological data is accumulating faster than people's capacity to analyze them, and; there are not enough bioinformaticians[15]. The National Science Foundation (NSF) identified bioengineering and bioinformatics as essential interdisciplinary for physical and life sciences and should be utilized to improve undergraduate research[16]. To respond to this changing landscape, the HGSC-G/GREAT focused efforts on identifying more program participants that would be interested in learning about bioinformatics and training them to increase their computational knowledge.

When we began the internship program, research mentors explained that the summer students did not need programming experience for the bioinformatics projects. During the 2008-2010 HGSC-G/GREAT summer program, we placed UR students in bioinformatics laboratories, and the success of these interns was cold to tepid. The interns that did not want to continue in the assigned primary project were given wet lab experiences to finish out their summer internship. Mainly 2 of the five students in this time frame gave it the 'old college try', and they turned out to be males (40%). We ascertained the reason the interns were reluctant to work in these projects. The students felt they needed a computer programming course and training before they would be comfortable working on a bioinformatics project. Because of this experience, we focused efforts on providing bioinformatics training opportunities to eliminate students' fears of programming and increase their willingness to train in this research area.

Before placing students in future bioinformatics laboratories, the program leadership decided to introduce bioinformatics to interns by developing a mini training course. Using *Backward Design* (UdD) to create a curriculum that would encourage the UR students in bioinformatics again, we developed a Bioinformatics BootCamp course and offered it to our trainees to address these concerns and increase their computational knowledge. Figure 1 is the UdD model of the HGSC-G/GREAT Bioinformatics Bootcamp course. There were certain elements (biomedical concepts, hands-on computational assignments) necessary for this course to be successful. This course included three types of students: (1) non-experienced students (with coding), (2) non-biology majors (computer science, physics, statistics), and the highly motivated students (self-taught coders). The mini course had two components to address the needs of all of these students. The biomedical topics portion included introducing DNA concepts, including Human Genome Project outcomes. The same instructor has taught this portion since the course inception. The bioinformatics portion has been taught by three different instructors using different coding programs.

We surveyed the instructors to determine the type of programming they used to teach the course (Figure 3) and based on this experience their opinion about the non-experienced student (Figure. 4). Each coding program had its benefits and was acceptable to the students in the course (Figure 2). In 2014, we planned to alter the questions with the change in instructor and did not give the full course survey to this cohort. The 2014 group thought the course was highly beneficial (4.7 mean on a 5-point Likert scale). Figure 2 shows that 50% of the 2014 group had experienced coders and coupled with Instructor B's observations that a few of these students were able to use what they learned can explain the high rating. At the end of each class, we had wrap-up sessions in addition to the formal evaluations; we learned that the bioinformatics portion was too fast paced and gave students too many coding challenges. The students gave this feedback in both the 2013 and 2014 summer program wrap up sessions, and interestingly these two groups had the highest number of students with bioinformatics experience. Based on this feedback, we attempted to solve the "coding overload" problem in 2015. We reasoned that we had not found the best approach that would also increase student proficiency in bioinformatics. With this understanding, the third instructor's strategy was to introduce a text-based interface that takes up far fewer computing resources than a full graphical interface. This knowledge allowed the instructor to take a simple approach to *demystify the command line*. Using the Unix command line approach went over well with the summer students because they could figure out how to use the information, actually accomplish the goal, and knew how they did it. We saw more students completing computational assignments in this group. In Figure 5, the students *agreed or strongly agreed* (100%) that the Bioinformatics Bootcamp course was conducted in a manner that served the student's needs. This teaching method inspired several of the students to continue working on the lessons after the course ended. In the first two years, when the course ended, the students had a sigh of relief and reported that they were more confused than before. Those students who had prior research experience or taught themselves how to code fared much better than those not exposed to bioinformatics concepts.

When we analyzed all of the surveys along with the course wrap up sessions, we determined the best approach for the mini course. Instructor C's teaching method for the Bioinformatics course was selected for subsequent years. This teaching method provided the most benefit to the novice student. We also found additional ways to use all of the instructors' approaches by incorporating them into the other G/GREAT program courses and training. We used Instructor A's approach in our Pre-Internship Course to provide G/GREAT interns exposure to bioinformatics before beginning the summer program. We used the online Codeacademy.com program to introduce the interns to Python. The more experienced students were both self-taught and had taken college courses. Their motivation to pursue this training prior to this program generated the idea to create self-paced training modules to increase their coding skills. We developed PERL training videos for use by the more experienced students during the course if they need more of a challenge and to continue learning beyond the course.

Two MARC (Minority Access to Research Careers) programs funded by NIH provided an intensive bioinformatics training for students from Minority Serving Institutions (MSI). The California State University-Los Angeles (CSULA) operated a summer bioinformatics education program emphasizing didactic preparation, research training, and professional development for upper-division undergraduates along with first and second-year graduate students[17]. The ten-week program consisted of 3 weeks of didactic instruction in bioinformatics, molecular life science, computer science, bioethics, and mathematics. This time was used to train in computer programming (initially C++, changed to Python) along with other bioinformatics tools like sequence alignment and microarray analysis. The rest of the seven weeks, the students performed bioinformatics research. Their data indicated that 89% of program graduates were in a career trajectory that would use bioinformatics[17]. When this article was published, eleven of their program graduates were still in undergraduate programs. The Pittsburgh Supercomputing Center (PSC) developed a training program to build bioinformatics at five Minority Serving Institutions (MSI). The PSC included a two-week-long Summer Bioinformatics Institute at PSC (sequence analysis, phylogenetics of proteins, Galaxy interface, and introduction to Next-Generation sequence bioinformatics). After the first two weeks, the interns conduct bioinformatics research for two months with a bioinformatics expert that result in completing conference-quality presentations. The program supported undergraduates, master's, and doctoral students. When asked about computational skills, there was a significant number of participants (database (73%), Python (68%), programming (68%), UNIX (77%)) that rated as a novice in the program pre-survey[18]. In contrast to the HGSC-G/GREAT program, both the CSULA and PSC bioinformatics programs provided more time with bioinformatics skill preparation and trained all of the participants in hands-on bioinformatics research projects. This allowed these programs to successfully place more bioinformatics trainees in the workforce. Our program is increasing the number of biomedical workforce professionals with bioinformatics tools with a smaller subset of them trained as experts in the field. The HGSC-G/GREAT Bioinformatics Bootcamp course is fulfilling a need as indicated by these programs. Feedback from PSC alumni currently employed in bioinformatics research careers stated that they wished they had learned

UNIX and run bioinformatics tools from the command line[18].

Published articles describing international bioinformatics internships was scarce. One example, in 2009, the Internship Program was organized by The Student Council of the International Society for Computational Biology[19]. The program aimed to help students from developing nations gain research and academic skills in computational biology. The data indicated that these participants were from Brazil, Estonia, India, and Kenya, and the status of the current students ranged from bioinformatician, researcher, research fellow, master student, Ph.D. student to Assistant Professor[19]. The alumni have had success in bioinformatics based on the outcomes reported. The background of the participants, college classification, major, or prior programming experience was not reported.

Many methods have been utilized to enhance bioinformatics in STEM that have resulted in an increase of trainees with this skill. Developing conferences (https://h3africa.org/; www.iscb.org,), courses (biohpc.learn.in.th/)[20,21], programs[21,23], summer internships[17-19], and workshops are avenues used to establish training in this area. Countries like Africa[24-27] and Thailand[28] have made tremendous efforts to increase bioinformatics capacity in their nations. For example, since 1996, efforts to increase bioinformatics infrastructure in Africa has led to the recent development of Human Heredity and Health in Africa initiative (www.h3africa.org). The Thai government realized the importance of this field and created a national policy to increase Thailand's participation in bioinformatics and genomics[28]. An example of this focus, the International Conference on Bioinformatics: North South Networking, by BIOTEC in collaboration with the Asia-Pacific Bioinformatics Network, was organized to promote awareness of bioinformatics in Thailand[28].

There was a dearth of research project data on medical student internships in bioinformatics. More studies are necessary to assess medical students' ability to train in bioinformatics during the first four years of medical school. One study, Dong et al. (2012) looked at the self-reported research experiences of medical students and the association with their performance in medical school and internship. Although this research study was not directly related, the outstanding ability to collect data from this target group was encouraging for other medical programs. They were able to collect data across multiple years. Between 1993 and 1999,1,112 medical students who graduated from the Uniformed Services University (USU) participated in the study[29]. Dong et al. (2012) collected 943 (85%) survey responses. The outcome measures were gathered across the medical education continuum from medical school preclinical study to clinical study and on into internship[29]. If medical programs would ascertain the interest of their graduates, then the success medical students have after they seek out bioinformatics research training opportunities could be tracked to determine feasibility and persistence in this field. Another study demonstrated that medical students could learn in an accelerated skills development course[30]. Zeng, Woodhouse, and Brunt (2010) showed that preclinical background and clerkship experiences impacted skills performance in an accelerated internship preparation course for senior medical students. Although this was a course for students interested in surgery, offering an internship in the senior year to medical students with some prior coding experience could increase their confidence to seek research training experiences while developing their clinical specialties.

Another concern facing clinicians is genetics in medicine has become a challenge for medical training. Bioinformatics efforts that appear to be wholly geared toward basic science are likely to become relevant to clinical informatics in the coming decade[31]. Tambi et al. (2018) designed a structured lesson plan to teach the specifics of bioinformatics for undergraduate medical students and was the first that outlined an effective dissemination strategy for integrating introductory bioinformatics into a medical curriculum. This study determined that introductory bioinformatics in a typical medical curriculum should focus on the following: comparison of nucleotide sequences and prediction of how a mutation affects the structure-function of the protein it translates[21]. They reported that 85% of the student cohort expressed confidence in being able to apply bioinformatics in their research projects[21]. Because of the advances genetics and genomics have made the NIH, CDC, and the Health Services and Resources Administration convened a workshop to identify practical strategies to educate primary care physicians involved in maternal and child health[32]. A significant challenge for this group, newborn screening, is a public health service available to all newborns. The advancements in genetic screenings will, in large part, require the primary care physicians to increase their understanding of genetics and genomics to help the patient interpret findings that may affect their care throughout their lifetime. This means that physicians, medical students, and clinical researchers need to learn the fundamentals of bioinformatics[21]. When this becomes standard curricula, we will also be able to increase medical students seeking opportunities to learn bioinformatics and therefore see an increase in physician-scientists in the biomedical workforce.

5.0 Conclusion

This study provides evidence that novice bioinformatics students can be trained to increase their interest, self-confidence, and proficiency when using the right computer language for the situation. After investigating the programming language choice for use to train the non-experienced student, the authors concluded that all approaches were important if used in the proper sequence of the summer research training program. When searching for research training opportunities for clinicians, we did not identify bioinformatics internships for medical students. The published programs we did locate did not highlight a focus on training future clinical researchers. Data was not provided that would illuminate the clinical interest of their program graduates. With the increased advancements in genetics in medicine, we potentially could see more medical students seek bioinformatics training opportunities as a result of changes in the medical school curriculum. The bioinformatics bootcamp approach can be generated at biomedical institutions around the country to provide introductory skills to clinicians. With the many problems observed in the clinic, these health care providers could use bioinformatics to solve many clinical issues. The same logic could be applied to health disparities and the use of bioinformatics.

6.0 Acknowledgements

Funding provided by the National Human Genome Research Institute/NIH R25HG006674.

Note: The corresponding author offers access to the surveys used in this study, on request.

References

1. National Institutes of Health. Physician-Scientist Workforce Working Group Report. 2014. https://acd.od.nih.gov/documents/reports/PSW_Report_ACD_06042014.pdf. Access February 7, 2020.
2. AAMC. Is an MD-PhD Right for Me? https://students-residents.aamc.org/choosing-medical-career/article/md-phd-right-for-me/. Access March 7, 2020.
3. Kaiser, J. NIH report warns of looming shortage of physician-scientist. Science. https://www.sciencemag.org/news/2014/06/nih-report-warns-looming-shortage-physician-scientists#. Published June 9, 2014. Accessed March 15, 2020.
4. Shaddox, C. Is the physician-scientist an endangered species? https://medicine.yale.edu/news/yale-medicine-magazine/is-the-physicianscientist-an-endangered-species/. 2011 Autumn. Access March 16, 2020.
5. Adams, D. Science. NIH report warns of looming shortage of physician-scientist. Science https://www.sciencemag.org/careers/2014/06/nih-report-warns-looming-physician-scientist-shortage Published June 10, 2014. Accessed March 15, 2020.
6. National Human Genome Research Institute. Educational Resources. Glossary of Genetics Terms. https://www.genome.gov/genetics-glossary/Bioinformatics. Accessed March 15, 2020.
7. Kemper, AR, Trotter, TL, Lloyd-Puryear, MA, Kyler, P, Feero, WG, Howell, RR. A blueprint for maternal and child health primary care physician education in medical genetics and genomics medicine: Recommendations of the United State Secretary of Health and Human Services Advisory Committee on Heritable Disorders in newborns and Children. Genetics in Medicine. 2009; 12 (2): 77-80.
8. Learning to Code - for Free.www.codeacademy.com. access March 7, 2020.
9. Build Skills with online courses from top institutions. Coursera. www.coursesera.ogr.Access March 7, 2020.
10. Author (2014). An exploratory assessment: developing pathways for underrepresented minorities into genomic science. Sage Open. 2014 October-December: 1–11 doi: 10.1177/2158244014560544
11. Wiggins, G. and McTighe,J. *Understanding by Design*. Association for Supervision and Curriculum Development (ASCD) 2005 2nd edition Alexandria, VA Paperback ISBN: 1-4166-0035-3
12. Manolio, T., Brooks, L., and Collins, F. A HAPMAP harvest of insights into the genetics of common disease. *J Clin Invest.* 2008.;118(5):1590-1605. https://doi.org/10.1172/JCI34772
13. National Institutes of Health All of Us Research Program. The future of health begins with you. https://allofus.nih.gov/
14. Sudmant, P., Rausch, T., Gardner, E., ….Korbel, J. An integrated map of structural variation in 2,504 human genomes Nature. 2015. vol 526: 75-81. doi:10.1038/nature15394
15. Chang, J. Core services: reward bioinformaticians. Nature. 2015. 520 (7546), 151-152. doi:10.1038/520151a
16. Revolutionizing Science and Engineering Through Cyberinfrastructure: Report on the National Science Foundation Blue Ribbon Advisory Panel on Cyberinfrastructure (whitepaper). 2003 Advisory *Panel on Cyberinfrastructure. www.cise.nsf.gov/sci/reports/atkins.pdf*

17. Krilowicz B., Johnston, W., Sharp, S., Warter-Perez, N., and Momand, J. A summer program designed to educate college students for careers in bioinformatics. CBE-Life Sciences Education 2007; 6: (74-83). doi:10.1187/cbe.06-03-0150
18. Mendez, R., Torres, J., Ishwad, P., Nicholas, H., and Ropelewski, A. Assisting bioinformatics programs at minority institutions: needs assessment, and lessons learned-a look at an internship program. http://dx.doi.org/10.1145/2949550.2949641
19. Anupama, J., Francescatto, M., Rahman, F., Fatima, N., DeBiasio, D., Shanmugam, A., Satagopam, v., Santos, A., Kolekar, P., Michaut, M., and Guney, E. The ISCB Student Council Internship Program: Expanding computational biology capacity worldwide. PLos Comput Biol 2018; 14 (1): e1005802. https://doi.org/10.1371/journal.pcbi.1005802
20. Wang, T. Course-based undergraduate research experiences in molecular biosciences-patterns, trends, and faculty support. FEMS Microbiology Letters 2017;364 (15): 1-9. Doi: 10.1093/femsle/fns157
21. Tambi, R., Bayoumi, R., Lansberg, P., Banerjee, Y. Blending Gagne's Instructional Model with Peyton's Approach to desgn an introductory bioinformatics lesson plan for medical students: proof-of-concept study. JMIR Med Educ 2018; 4 (2). doi: 10.2196/11122
22. Goode, E., and Trajkovski, G. Developing a truly interdisciplinary bioinformatics track: work in progress. JCSC 22, 6 (June 2007).
23. Hemminger, B., Losi, T., and Bauers, A., (2005) Survey of Bioinformatics programs in the United States. J American Society for Information Science and Technology, 56 (5): 529-537.
24. Shaffer, J., Mather, F., Wele, M., Li, J., Tangara, C., Kassogue, Y., Srivastav, S., Thiero, O., Diakite, M., Sangare, M., Dabitao, D., Toure, M., Djimde, A., Traore, S., Kiakite, B., Coulibaly, M., Liu, Y., Lacey, M., Lefante, J., Koita, O., Schieffelin, J., Krogstad, D., and Doumbia, S. Expanding Research Capacity in Sub-Saharan Africa Through Informatics, Bioinformatics, and Data Science Training Programs, in Mali. Frontiers in Genetics 2019; 10 (331): 1-13. doi: 10.3389/fgene.2019.00331
25. Ojo, O., and Omabe, M. Incorporating bioinformatics into biological science education in Nigeria: Prospects and challenges. Infection, Genetics and Evolution 2011;11: 784-787. doi:10.1016/j.meegid.2010.11.015
26. Fatumo, S., Adoga, M., Ojo, O., Oluwagbemi, O., Adeoye, T., Ewejobi, I., Adebiyi, M., Adebiyi, E., Bewaji, C., and Nashiru, O. Computational biology and bioinformatics in Nigeria. PLoS Computational Biology. 2014; 10 (4): e1003516. doi:10.1371/journal.pcbi.1003516
27. Bishop, O., Adebiyi, E., Alzohairy, Al, Everett, D., Ghedira, K., Ghouila, A., Kumuthini, J., Mulder, N., Panji, S., and Patterton, H. Bioinformatics Education-perspectives and challenges out of Africa. Briefings in Bioinformatics 2014; 16 (2): 355-364. doi: 10.1093/bib/bbu022
28. Tongsima, W., Tongsima, S., and Palittapongarnpim, P. Outlook on Thailand's Genomics and Computational biology research and development. PLoS Comput Biol 2008; 4(7): e100115. doi:10.1371/journal.pcbi.1000115
29. Dong, T., Durning, S., Gilliland, W., Waechter, D., Cruess, D., DeZee, K., Calloway, M., Artino, A. Exploring the

relationship between self-reported research experience and performance in medical school and internship. Military Medicine 2012; 177, 9:11.

30. Zeng, W., Woodhouse, R., and Brunt, M. Do preclinical background and clerkship experiences impact skills performance in an accelerated internship preparation course for senior medical students? Surgery 2010; 148 (4); 768-777. doi:10.1016/j.surg.2010.07.022
31. Altman, R. Bioinformatics in support of molecular medicine. Proceedings of the AMIA Symposium, 1998; 53–61. PMCID: PMC2232090
32. Kemper, A., Trotter, T., Lloyd-Puryear, M., Kyler, P., Feero, W., and Howell, R. A blueprint for maternal and child health primary care physician education in medical genetics and genomic medicine: Recommendations of the United States Secretary for Health and Human Services Advisory committee on heritable Disorders in Newborns and children. Genetics in Medicine 2010; 12(2): 77-80.