

Master of Science in Bioinformatics Program Review and Assessment Report 2021–2022

1. Program Summary

The massive biological data, such as data generated from the human genome project and electronic health records, has led to the great need of computational techniques and machine learning over the big data. The field of bioinformatics emerges when computing meets information. Simply speaking, bioinformatics is an interdisciplinary field among biology, computer science, and information technology that develops algorithms, software tools, and data repositories for aggregating, storing, curating, analyzing, and visualizing biological data. Bioinformatics has significantly contributed to the cures for human diseases, improvements of crop quality and production, creation of new technologies, and novel applications to medicine and industries. With the high throughput data generation in biological science, the need for bioinformaticians is on the rise. Hence, it is necessary to train the next generation of bioinformatics professionals with the required skills to prepare them as future successful bioinformaticians.

Bioinformatics programs have been offered at different levels in the US, from BS to MS and to PhD programs. Due to its multidisciplinary nature, it is important that the bioinformatics curriculum benefits both the computer scientists and the biologists. Many approaches and strategies for bioinformatics education have been proposed in various bioinformatics programs. Especially, there is great advocacy for conveying the computational and information technology skills for biologists to meet the challenges for the big data sciences. The biomedical data analytics, especially for the -omics era, is another challenging area. In addition, the program competencies have been discussed among researchers and educators. For example, some previous work concluded that the five broad competency areas required for bioinformaticians are biology, computer science, statistics, ethics, and core bioinformatics.

Bioinformatics is a cutting-edge graduate program in the School of Informatics and Computing (SoIC) at IUPUI, which is the first school of its kind in the US, focused on innovative research in fast-paced and dynamic informatics fields. The bioinformatics program is a relatively new interdisciplinary field and focuses on translational research that uses computational approaches to transform biological data into scientific discoveries that help us better understand the life science and improve patient care. No national ranking for such a program exists because of its degree of specialization.

2. Program Processes

2.1 Program contents and challenges

Extensive efforts have been taken by the Bioinformatics Program for improvement of the program. The emerging trend of large-scale biological data has restrained the traditional biologists from using the information and has led to the needs of programming and computer software for data analytics, which led to the ‘data-driven science.’ At the same time, the Bioinformatics Program has to adapt to meet these new needs. In addition, as an emerging interdisciplinary field, the Bioinformatics Program is facing new challenges for curriculum design, course organizations, and training delivery approaches.

The program challenges also affect the trainees ultimately. For example, it is natural for some trainees to wonder whether they have been provided the best practical training in bioinformatics so that they are career-ready upon completion of their degree programs. The Bioinformatics Program is expected to have trainees from different life science and computer science backgrounds, as the bioinformatics is the bridge between life science and computer science. The program is expected to deliver intensive training in both biology and computing skills to provide the workforce with individuals with an interdisciplinary set of skills necessary for bioinformatics projects.

Students with strong computer science background are expected to have the capability to create and modify bioinformatics software systems and/or applications to address questions in bioinformatics research. Students with life science backgrounds are expected to have applied knowledge on how to operate existing bioinformatics tools and to apply the tools for specific bioinformatics projects. Their strong life science background facilitates them in the interpretation of biological experimental results.

In addition, it is necessary to include hands-on training on online data repositories, such as UniProt and GenBank, and bioinformatics tools, such as the Basic Local Alignment Sequence Tool (BLAST) for identifying sequence similarities in both DNA and protein sequences. It would be better for a single application to integrate multiple bioinformatics tools for training the students in bioinformatics.

It is a general perspective that the main areas of bioinformatics are genomics, transcriptomics, proteomics, and systems biology. Genomics includes sequencing, assembly, and analysis of DNA, transcriptomics studies gene regulation and RNA sequences, proteomics specifically deals with protein structures and functions, and systems biology involves computational data modelling of biological systems, which can be molecule, tissues, organs, or whole organism.

2.2 Program Competencies

Existing program courses and curriculum designs are based on these four major areas of bioinformatics: genomics, transcriptomics, proteomics, and systems biology. The major competencies of bioinformatics programs are listed below:

- biological database management,
- genomics data analysis,
- transcriptomics data analysis,
- proteomics data analysis,
- statistical methods in bioinformatics,
- algorithms,
- bioinformatics programming,
- systems biology,
- structural biology,
- next generation sequencing,
- third generation sequencing,
- data mining,
- text mining,

- big data analysis,
- computer modeling,
- machine learning,
- and other major skills.

2.3 Structure, breadth, and depth of the curriculum

This program is structured and has a need to be taken in a specific order for some advanced courses. It is set up so that the program is built as blocks to help ensure the student is given its foundation to move onto the next section.

The required core courses give the basic overview of the bioinformatics program and provide hands-on experiences for the various techniques and tools used in bioinformatics. The advance core courses prepare students to further specialize in a certain area and keep up with the ever-changing emerging areas in bioinformatics. The elective courses provide the student with a wider breath to learn of related fields and reinforce their knowledge in bioinformatics. Basically, the curriculum design ensures that, no matter what a student background is, the student can find courses in bioinformatics that suitable for their understanding of bioinformatics.

The Bioinformatics program is diverse so each class must meet a specific measure to ensure the student has a foundation on which to build.

2.4 Interdisciplinary program

Bioinformatics is an interdisciplinary program in the areas of computational biology, computer sciences, and information technology. This gives the students more areas of expertise to choose from. For example, for the biology-oriented students with limited background, the programming courses (Shell script, R, and Python) are taught in their first semester entering the programs. This makes sure they are ready to work on bioinformatics projects from the beginning. Students with strong computer science background have the opportunities to learn more on biological science and computational bioinformatics.

2.5 How has the program curriculum responded to new directions in the discipline

The curriculum has been adjusted to meet new directions since the program started and will continue to be adjusted because bioinformatics is an area that is continuously evolving. For example, over the past few years, research in single cell sequencing, nanopore sequencing and ATAC-Seq have become increasingly active. Faculty in the bioinformatics program have continuously added these new topics to the core courses, such as INFO-B 519 Introduction to Bioinformatics and INFO-B 528 Computational Methods for Analyzing High-Throughput Biological Data. The program continues to adapt itself to the emerging new technologies. Our curricula and syllabus are kept up-to-date to meet the market needs.

A curriculum committee has been formed in the department to go over the curricula and syllabi periodically, making sure they are updated.

3. Program Learning Outcomes

Principles of Graduate and Professional Learning (PGPL)

- | | |
|--|--------------------------|
| 1. Knowledge and skills mastery | <i>Major emphasis</i> |
| 2. Critical thinking and good judgment | <i>Moderate emphasis</i> |
| 3. Effective communication | <i>Some emphasis</i> |
| 4. Ethical behavior | |

Assessment methods

- A Assignments
 Q Quiz
 M Midterm quiz/exam
 E End-term quiz/exam
 F Final project
 P Presentation

The expected program learning outcomes are based on the major competencies and major bioinformatics areas. Upon completion of the graduate degree program, a student is expected to possess the following bioinformatics capabilities.

Learning outcomes	Course	RBT	PGPL	Assessment
1. Analyze biological data set with state-of-the-art computational tools and methods	INFO-B 519	3	1	AME
2. Design software pipelines for automatic analysis of big biological data and to find biological information from the data	INFO-B 519	4	1 2	AME
3. Identify and interpret large size health data with missing values	INFO-B 518	3	1,2	AQMFP
4. Infer and justify small size health data specific to diseases	INFO-B 518	3	1,2	AQMFP
5. Correlate massive phenotypic and genotypic data	INFO-B 518	3	1,2	AQMFP
6. Decide and model population, sampling and hypothesis testing for specific diseases	INFO-B 518	4	1,2	AQMFP
7. Design and formulate sampling and hypothesis testing for hospital data and Insurance data set to evaluate the complexities	INFO-B 518	4	1,2	AQMFP
8. Select and generate regression analysis and other statistical analysis for precision medicine applications	INFO-B 518	4,5	1,2,3	AQMFP
9. Outline and formulate paper presentation	INFO-B 518	4,5	1,2,3,4	P
10. Construct and rearrange project design, writing, analysis, and presentation	INFO-B 518	5,6	1,2,3,4	QMFP
11. Develop and revise programs to perform data analytics on large, complex datasets in R	INFO-B 518	4,5,6	1,2	AQMFP
12. Analyze and process microarray datasets and functionally interpret the results in light of molecular biology.	INFO-B 528	4	1, 2	AME

13. Analyze time course RNA and protein expression levels and model of expression data.	INFO-B 528	4	1	AME
14. Analyze genomes comparatively and functionally Predict operon structure Methods for function prediction Studying evolution of operon structure Apply the principles to metagenomic context	INFO-B 528	4	1	AME
15. Analyze environmental microbial genomic data, resources available for metagenomics, metatranscriptomics, operons and transcription units taxonomic mapping, microbial abundance, interactions, and pathways.	INFO-B 528	4	2	AME
16. Evaluate prediction algorithms and their applications in understanding regulatory systems biology (using representations of regulatory motifs).	INFO-B 528	5	2	AM
17. Analyze networks by applying a range of algorithms.	INFO-B 528	4	2	AME
18. Evaluate biological networks, by developing and applying computational approaches for analyzing regulatory, protein-protein, genetic, and chromosomal interaction mapping data.	INFO-B 528	5, 6	1	AME
19. Evaluate current approaches for determining the structure, dynamics, and evolution of biological networks.	INFO-B 528	5	4	AME
20. Understand basic concepts and theories in machine learning.	INFO-B 529	2	1	AME
21. Use machine learning algorithms and tools in biological data analysis.	INFO-B 529	3	1	AMEF
22. Understand the pros and cons of various machine learning algorithms in applications.	INFO-B 529	2	1	AME
23. Select correct machine learning tools and parameters in biological data analysis and to find valuable information from the data.	INFO-B 529	4	2	AMEF
24. Evaluate common problems in biomedical informatics, such as sequence alignment, genome arrangement, and peptide identification	INFO-B 536	5	1	AME
25. Analyze time and space complexity and other theoretical concepts used in algorithm analysis and complexity analysis	INFO-B 536	4	1	AME
26. Apply abstract data structures to solve problems in biomedical informatics	INFO-B 536	3	1, 2	AMEF
27. Compare the pros and cons of computational methods for a biomedical problem and choose appropriate methods	INFO-B 536	4, 5	1, 2	AMEF

28. Evaluate the similarity between new problems and existing problems and adapt computational methods designed for existing problems to new problems	INFO-B 536	5, 6	1, 2	AMEF
29. Design computational methods using a greedy, brute-force, divide-and-conquer, or dynamic programming approach	INFO-B 536	6	2	AF
30. Evaluate biomedical problems using example-based problem-solving skills and iterative refinement skills	INFO-B 536	5	2	AMEF
31. Design and perform experiments for evaluating computational methods and publicly present experimental results	INFO-B 536	3, 6	1, 2, 3	F
32. Learn basics of operating systems and running shell commands to perform routine tasks in Linux.	INFO-B 573	4	2	AME
33. Learn Python Basics and able to script for analyzing simple to complex biological data sets with advanced integrated aspect of Linx and python	INFO-B 573	3	1	AME
34. Analyze complex biological data using advanced Python skills with PANDAS and NUMPY in Linux systems.	INFO-B 573	4	1	AME
35. Learn R programming basics, analyze read, edit and write a file in R, perform statistical testing and generate graphics. Students will be able to write R scripts for analyzing biological data sets.	INFO-B 573	4	2	AME
36. Learn to write HTML code and develop personalized website using content manage systems.	INFO-B 573	5	2	AME
37. Understand dealing with data and database management systems. Students will also learn different major public databases in biological domain.	INFO-B 573	4	2	AME
38. Learn various bioinformatics applications- command line and interface	INFO-B 573	5	1	AME
39. Write a report and give an oral presentation grounded in an appropriate review of the literature.	INFO-B 573	6	3	F P
40. Be introduced to the basics of molecular biology and need of programming in biology.	INFO-B 636	4	2	AME
41. Learn basics of operating systems and running shell commands to perform routine tasks.	INFO-B 636	3	1	AME
42. Learn Python Basics and able to script for analyzing simple to complex biological data sets with advanced integrated aspect of unix and python.	INFO-B 636	4	1	AME
43. Learn R programming basics, analyze read, edit and write a file in R, perform statistical testing and	INFO-B 636	4	2	AME

generate graphics. Student will be able to write R scripts for analyzing biological data sets.				
44. Learn writing html code and developing code to connect databases with Python/PHP.	INFO-B 636	5	2	AME
45. Understand dealing with data and database management systems. Student will also learn different major public databases in biological domain.	INFO-B 636	4	2	AME
46. Learn various bioinformatics applications- command line and interface	INFO-B 636	5	1	AME
47. Write a report and give an oral presentation grounded in an appropriate review of the literature.	INFO-B 636	6	3	FP

RBT: Revised Bloom's Taxonomy

4. Program Assessment

4.1 Direct measures






4.1.1 Methodology









To evaluate how well students are meeting the program's intended learning outcomes (LOs), we conducted quantitative analysis of student grades for all assignments from three core courses in the program. Anonymized student grades were collected from instructors for courses taught in Spring 2022 and Fall 2022. The grades were broken down by individual assignment, project, exam, or quiz. Grades were renormalized so that they range from 0 to 100 points to allow meaningful comparison between courses. We then asked course instructors to map each assignment to program-level LOs. The collective number of course enrollments represented in this sample is 79. Note that this number does not represent unique students, because students typically enroll in multiple courses simultaneously. The collected dataset covers 15 of the program's 47 learning outcomes.

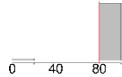
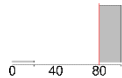
4.1.2 Results

We computed grade distributions from the above data. These distributions represent instructor assessment of students' mastery of each program-level LO. The distributions are plotted as histograms in the following table, broken down by competency area and then by course. The red line within the histogram plot represents an 80-point grade, the program's threshold for passing. We also include the percentage of students (shown to the right of the histogram) who are considered to have mastered the intended outcome, having achieved a passing grade (≥ 80 points). The percentile is color coded in green ($\geq 70\%$), orange ($\geq 50\%$), or red ($< 50\%$).

Analysis of student grades reveals that all LOs show evidence of good student achievements, with over 80% of students obtaining a passing grade in exams, assignments, and/or projects. Additionally, 0 LOs demonstrate moderate or low success (with less than 80% or 50% of students achieving a passing grade).

INFO B573	Grade distribution	% Students meeting learning outcome
LO1: Learn basics of operating systems and running shell commands to perform routine tasks in Linux.		95.5%
LO2: Learn Python Basics and able to script for analyzing simple to complex biological data sets with advanced integrated aspect of Linux and python.		100%
LO3: Analyze complex biological data using advanced Python skills with PANDAS and NUMPY in Linux systems.		100%
LO4: Learn R programming basics, analyze read, edit and write a file in R, perform statistical testing and generate graphics. Students will be able to write R scripts for analyzing biological data sets.		100%
LO5: Learn to write HTML code and develop personalized website using content manage systems.		100%

INFO B528	Grade distribution	% Students meeting learning outcome
LO1: Analyze and process microarray datasets and functionally interpret the results in light of molecular biology.		100%
LO2: Analyze time course RNA and protein expression levels and model of expression data.		85.7%
LO3: Analyze genomes comparatively and functionally.		95.2%
LO4: Analyze environmental microbial genomic data, resources available for metagenomics, metatranscriptomics, operons and transcription units taxonomic mapping, microbial abundance, interactions, and pathways.		95.2%
LO5: Evaluate prediction algorithms and their applications in understanding regulatory systems biology (using representations of regulatory motifs).		90.5%
LO6: Analyze networks by applying a range of algorithms		95.2%
LO7: Evaluate biological networks, by developing and applying computational approaches for analyzing regulatory, protein-protein, genetic, and chromosomal interaction mapping data		95.2%
LO8: Evaluate current approaches for determining the structure, dynamics, and evolution of biological networks		95.2%

INFO B519	Grade distribution	% Students meeting learning outcome
LO1: To familiarize students with some of the basic computational problems in bioinformatics		97.1%
LO2: To familiarize students with basic methods and tools for solving computational problems in bioinformatics		97.1%

4.2 Program assessment with indirect measures

The various bioinformatics programs have been studied regarding their hosting departments and schools in the United States. One of the major resources is the websites of the corresponding Bioinformatics programs, including the program outcomes, students, faculty, and other information. The second source is the US bioinformatics education provided by the Bioinformatics Organization which lists the bioinformatics programs:

http://bioinformatics.org/wiki/Education_in_the_United_States

It arranges all the bioinformatics programs in the US, for each state and each university. There are total 68 Bioinformatics Programs listed from this web site, including BS, MS, and PhD degree programs and certificate programs.

Career for Bioinformatics students

The job market for bioinformatics students is promising. The bioinformatics graduates can work on various related fields with varying titles. Sample industrial job titles include data scientist, research associate, computational biologist, bioinformatics scientist, bioinformatics analyst, software engineer, senior software engineer, and others. There are also openings in academia as postdoctoral fellows, research scientist, research professor, and faculty positions (such as assistant professors for Ph.D. students in Bioinformatics). Below are the market salary information for a few example positions.

- Bioinformatics scientist:** According to PayScale, the average salary for a Bioinformatics scientist is \$99,555 per year, with a reported salary ranges from \$67,000 to \$100,000 per year. The highest paid Bioinformatics scientist live in San Diego and Boston, according to data from PayScale, where the highest reported salaries are \$142,000 and \$164,000 per year.
- Bioinformatics Analyst:** According to PayScale, the average salary for a Bioinformatics analyst is \$79,394 per year, with an average reported salary of \$50,000 to \$82,500 per year. The highest paid Bioinformatics analyst resides in foster city according to data from PayScale, where the highest reported salary is \$110,000 per year.
- Bioinformatics software engineer:** According to PayScale, the average salary for a Bioinformatics software engineer is \$90,000 per year, with an average reported salary range of \$85,000 to

\$135,000. The highest paid Bioinformatics software engineer resides in San Francisco, according to PayScale data, where the highest reported salary is \$160,000 and \$150,000 per year.

- **Bioinformatics Developer:** According to PayScale, the average salary for a Bioinformatics developer is \$83,190 per year, with a reported salary range of \$58,000 to \$85,000 per year. The highest paid Bioinformatics Developer live in San Diego according to data from PayScale, where the highest reported salary is \$112,000
- **Bioinformatician:** According to PayScale, the average salary for a Bioinformatician is \$77,997 per year, with a reported salary range from \$53,000 to \$78,000 per year. The highest paid Bioinformatician, according to data from PayScale, resides in Houston with reported highest salaries of \$119,000 per year, respectively.

Skill requirements

Detailed analysis of the skill requirements for different job openings has also been performed. Based on indeed.com, the most required skills for bioinformatics jobs are programming skills, bioinformatics algorithms, data mining, statistical analysis, database managements, genomics, next generation sequencing, big data analytics, bioinformatics software tools, and others. More information of the required skills in bioinformatics is summarized below:

- **Bioinformatics tools:** Samples tools are sequence alignment tools, such as BWA, the Genome Analysis Toolkit (GATK).
- **Statistical software systems:** such as SAS. In addition, statistical analysis using R or Python is highly demanded.
- **Programming skills:** It was required for most bioinformatics jobs to be familiar with programming languages. Other than general programing, for bioinformatics, the most popular programming languages or script languages are Perl, Python, C++, and Matlab. At least one of these programming languages is listed in the job requirements.
- **Biology knowledge:** Knowledge from molecular biology, cancer biology, and/or modern biology is needed. This is the domain knowledge needed to understand and analyze the data, and interpret the analytical results.
- **Genomics and genetics:** Genomics and genetics are the core bioinformatics skill sets and are required for almost every bioinformatics job. For example, skills for high throughput sequencing technologies, next generation sequencing, and computational genomics are in high demands.
- **Database management:** Databases, including traditional relational database systems (e.g., SQL) and other big data databases (e.g., TCGA) are common required knowledge for potential bioinformatics job seekers. Especially, for big data management or analysis using cloud, such as AWS, it is a widely open field and is in great demands.
- **Machine learning:** for example, hierarchical clustering and decision trees are common required techniques.

- Some additional skills required for the jobs are good communication skills and presentation skills.

It can also be observed that there are some overlaps between the competencies and the skill requirements. In fact, the more overlap, the better, as it shows that our program competencies meet the job market for bioinformatics.

4.3 The MS bioinformatics employment data

The employment rate for MS in bioinformatics program at IUPUI is 100%, 74%, 100% and 94% for 2019, 2020, 2021 and 2022 respectively. Low employment in 2020 are likely due to the impact of COVID. For Bioinformatics MS graduates, the average salary was \$71,000, \$97,500, \$65,000 and \$65,000 for 2019, 2020, 2021 and 2022, respectively (Fig. 6). The common position title and the company names are listed below.

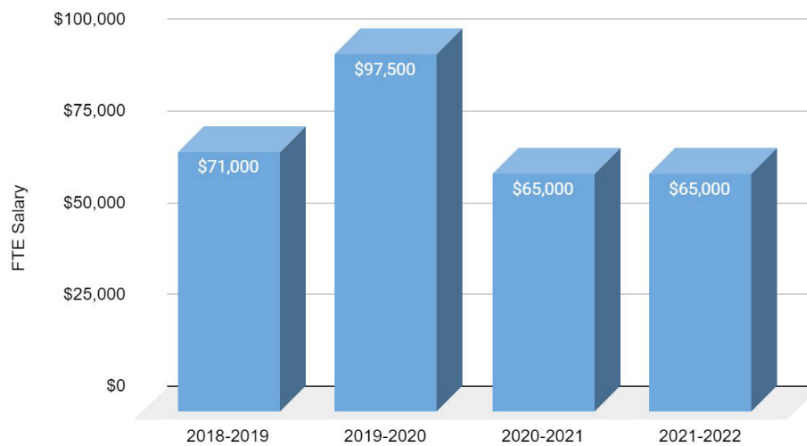


Figure 6. Average salary of Bioinformatics MS graduates

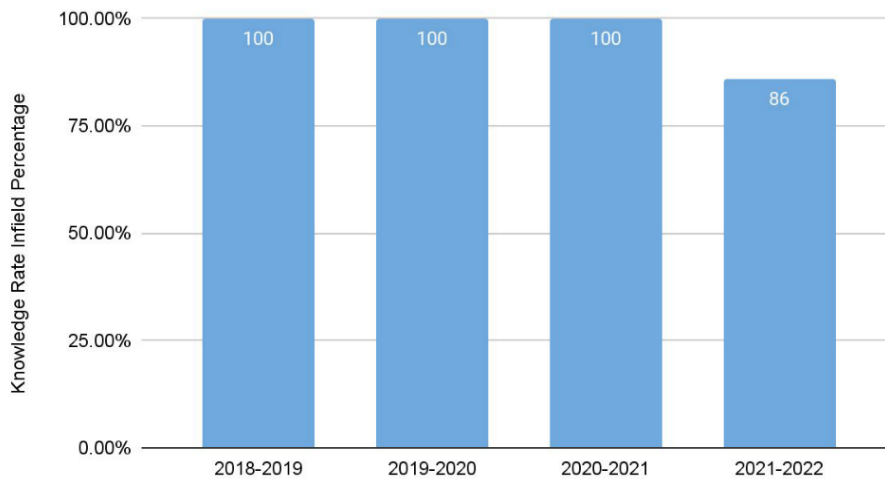


Figure 7. Average knowledge rate infield of Bioinformatics MS

Example Position Titles for Bioinformatics MS Graduates for 2019-2022:

- Bioinformatician
- Bioinformatics Engineer
- Technology Analyst
- Associate Bioinformatics Scientist
- Data Engineer

Company Lists for Bioinformatics MS Graduates for 2019-2022:

- Eli Lilly
- Personalis
- Covance
- Indiana University School of Medicine
- Mayo Clinic
- Washington University of Medicine
- Columbia University

4.4 Teaching Assessment

As bioinformatics is a practical subject, the teaching methods are a combination of lectures, problem-solving exercises, laboratory sessions, seminars and workshops, small group course projects, theses, internships, and independent studies.

How do we measure each of the desired behaviors listed as assessment methods?

- The assignments, quizzes, and exams are graded, and the scores are the measures.
- The projects are measured by the professors on the quality of the written reports, posters, and peer-reviewed publications.
- The thesis is evaluated by the thesis committee based on the quality of the thesis work, including the written report and the oral defense.
- The practical exercises, the internships, and the lab sessions are evaluated by the practical solutions, the skills the students demonstrated, and the student career opportunities.
- Retention and graduation rates are tracked and studied.

4.5 Enrollment Assessment

The Enrollment of the MS program in Bioinformatics students has been assessed with two metrics: the number of newly enrolled students

- **The total number of new MS students in the Bioinformatics program:** the last 5 years of MS students in the program has been illustrated in Figure 8, which shows the number of students was stable in general, with an exception in 2020 Fall due to the impact of COVID-19.

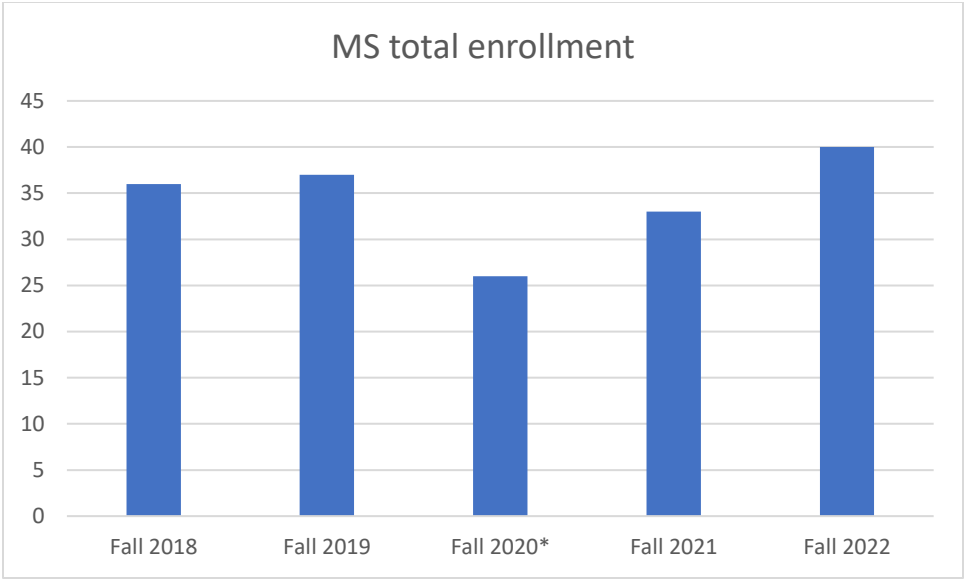


Figure 8. Numbers of newly enrolled students in the past 5 years.

- **The total credit hours of the Bioinformatics course enrollments:** for the last 5 years, the total credit hours of students enrolled in the Bioinformatics courses in the fall semester are illustrated in Figure 9. The same pattern as the headcount is observed that credit hours have been steadily increasing since COVID-19 and have recovered to the level of pre-COVID.

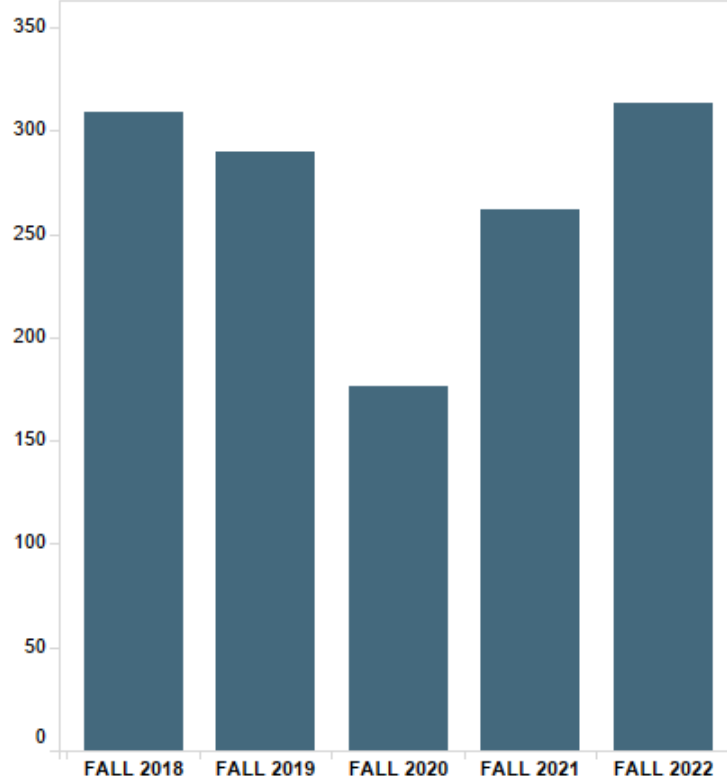


Figure 9. Fall semester credit hours of the Bioinformatics MS program in the past 5 years.

Combining these metrics of the enrollment assessment, the MS in Bioinformatics program has been healthy for the last few years. Even with the impact of COVID-19, the program student number and the credit hours have gradually bounced back.

5. Improvements of the Bioinformatics Program

The information learned from this study provides great guidance for the redesign of the curriculum, help for the students to self-evaluate their potential bioinformatics skills, and promotion of hands-on experience for bioinformatics professors, educators, and students.

The bioinformatics program at IUPUI offers both MS and PhD degrees. After revisiting all the curriculums and reevaluating the competencies, we will further improve the bioinformatics programs in the following matters:

- Eliminate redundant contents: By examining the program competencies and the individual course outcomes, the redundant contents across courses have been identified and resolved.
- Develop new course modules: to cover all the required skills, modular courses of 1 or 2- credits have been designed to cover specifically of one or two required skills. This guarantees that our students are career ready upon graduation. For example, for the biology-oriented students with limited background, the programming courses (Python, Shell scripts, and SQL) are taught in their first semester entering the programs. This makes sure they are ready to work on bioinformatics projects from the beginning.
- Reorganize the course offering schedule: The course offering schedule has been revisited to make sure that the students meet all the prerequisites when they are ready to learn a new skill.

In addition, all the students in our bioinformatics program are provided the opportunities to have hands-on experiences, starting from the beginning of their MS programs. All full-time MS students are provided with partial scholarships and hourly research assistantship. All the PhD students are fully supported. With acceptance of the scholarships, the students are expected to work on bioinformatics projects upon entering the programs. The student records showed that this is a great way to increase the students' experiences and to improve the portfolios of the students when they are ready for jobs.