IEEE
COMPUTER
SOCIETY

www.computer.org

Bioinformatics:
At the Intersection of Computer Science, Mathematics, and Biology

TABLE OF CONTENTS

INTRODUCTION

Bioinformatics is the study of storing, retrieving, and analyzing biological data. The interdisciplinary field combines computer science, biology, and mathematics into tools to improve our understanding of nature. Early bioinformatics applications were used to sequence DNA in the 1990s through the Human Genome Project; however, as large amounts of data become more central to biological research, bioinformatics is occupying a much broader space in the general field of science.

IEEE
COMPUTER
SOCIETY

LOOKING AHEAD

A Community of Computing Leaders

The IEEE Computer Society is the world's home for computer science, engineering, and technology. A global leader in providing access to computer science research, analysis, and information, the IEEE Computer Society offers a comprehensive array of unmatched products, services, and opportunities for individuals at all stages of their professional careers. Known as the premier organization that empowers the people who drive technology, the IEEE Computer Society offers international conferences, peer-reviewed publications, a unique digital library, and training programs.

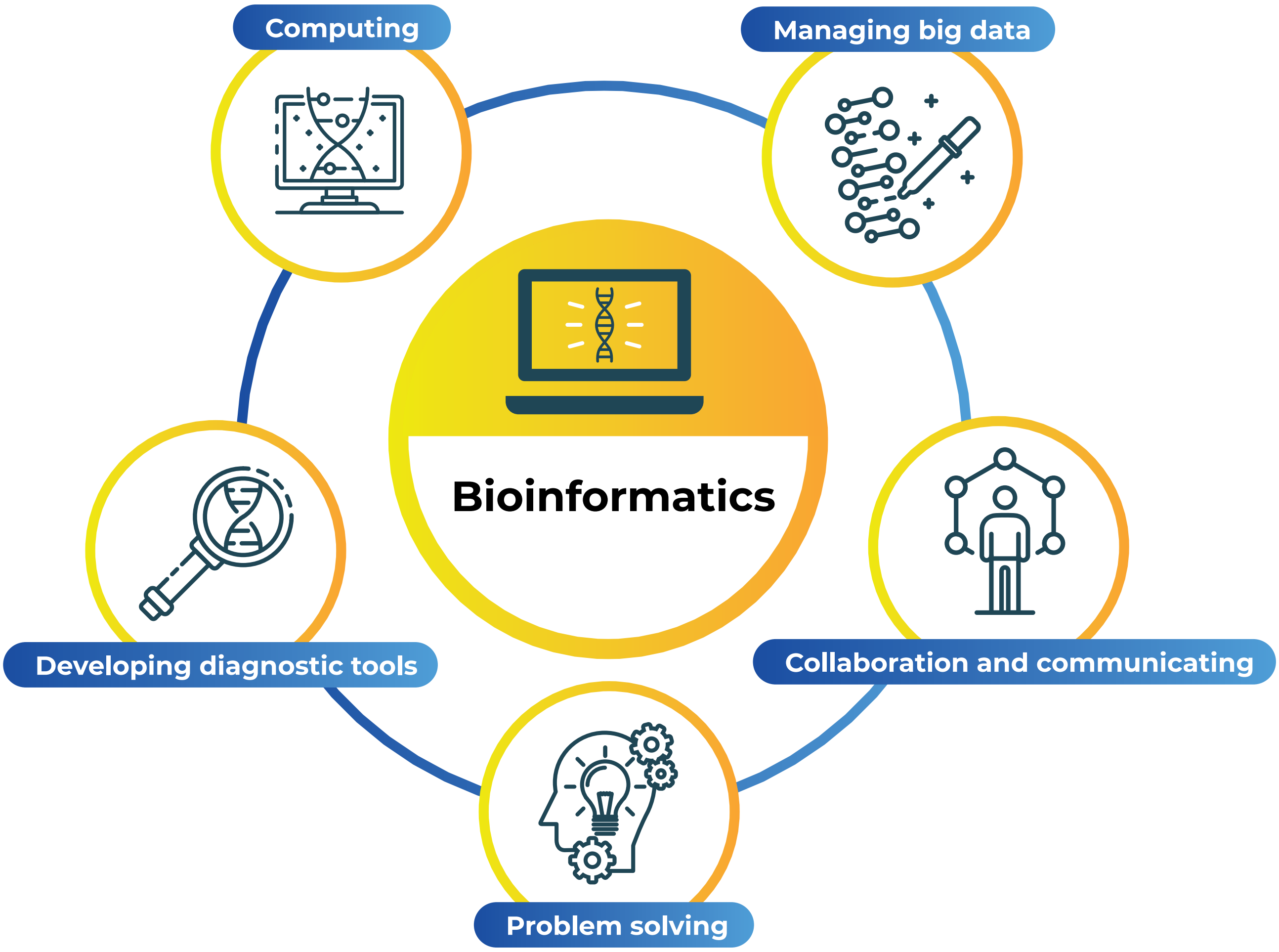
Visit computer.org to launch your career.

Bioinformatics now includes the creation and advancement of databases, algorithms, computational and statistical techniques, and theory to collect, manage, and analyze biological data. With the U.S. Bureau of Labor and Statistics forecasting a 22% growth rate in computer-based analysis jobs by 2030—a growth rate four times higher than the national average—“bioinformaticist” may become a more popular career path for scientists.

Below is a review of the bioinformatics field, as well as a look into the industry and careers for prospective bioinformaticists or individuals just wanting to know more about this increasingly diverse and ubiquitous profession.

With the U.S. Bureau of Labor and Statistics forecasting a 22% growth rate in computer-based analysis jobs by 2030—a growth rate four times higher than the national average—“bioinformaticist” may become a more popular career path for scientists.

What is Bioinformatics?



Source: NHS

Applications in Bioinformatics

Advances in biology, mathematics, and especially computer science have dramatically expanded the field of bioinformatics.

The immense amount of data created by modern biomedical instruments requires a massive amount of computing to translate into knowledge.

Bioinformaticists today are developing new analytical tools enabling researchers to perform large-scale integrative analysis of data from a myriad of sources in a rigorous and reproducible manner—for example, contact tracing for the COVID-19 pandemic and machine-learning-assisted analysis of digital pathology data.

There is no definitive list of bioinformatics applications, but a few of the more popular ones include gene and protein sequencing, structural bioinformatics, network and systems biology, and DNA computing.

Sequencing in bioinformatics includes the analysis of DNA, RNA, or protein data to better understand their features, purpose, configuration, and evolution.

Companies that offer human genome mapping, like 23andMe, rapidly sequence an individual's entire genome and offer their customers insights into their heritage, as well as the risk of some genetic diseases. The company also acts as a repository for massive databases of genetic data. 23andme boasts the world's largest research database of microarray genotype data in the world.

The CRISPR/Cas9 gene editing tool, a significant breakthrough in genetics, allows researchers to easily alter DNA sequences and modify gene function by precisely cutting DNA and then modifying, replacing, or deleting the genetic information. Its potential applications include correcting genetic defects and treating and preventing genetic diseases.

Structural bioinformatics concentrates on the analysis, prediction, and manipulation of the 3D structure of biological macromolecules, including DNA, RNA, and proteins. As form predicts function, this structural information is truly important. AlphaFold is an artificial intelligence system developed by DeepMind, in partnership with EMBL's European Bioinformatic Institute, which predicts a protein's 3D structure from its amino acid sequence. AlphaFold has reported an accuracy competitive with traditional experimental methods. Developers of the high-profile bioinformatics system recently created an open-source database filled with more than 350,000 3D protein structures predicted by AlphaFold.

Network and systems biology uses a bioinformatics approach to drug discovery by developing virtual models of entire physiological environments to test protein targets and the effects of modifying them without

Steady growth for the bioinformatics market, reaching about US$45.6 billion from 2022 to 2030—a compound annual growth rate of 16.3%.

losing the key molecular function. Applications in this bioinformatics niche are powerful tools for iteratively testing new therapeutics using math and computers instead of living animals and humans to predict interactions between drugs and to study disease comorbidity.

A team of researchers at the Wellcome Sanger Institute and ETH Zurich, Switzerland, recently reported using a high-throughput surface receptor screening method to systematically map the direct protein interactions across a recombinant library encompassing most of the surface proteins on human leukocytes. Simply put, they created a map of the human immune system. The connections uncovered by this map may bring about new therapeutic opportunities.

DNA computing is a bioinformatics application still in early development. With DNA computing, DNA, biochemistry, and molecular biology replace the electronic hardware and silicon chips of traditional computing. And unlike traditional computing, which uses a binary alphabet of zeros and ones, DNA computing has a four-character genetic alphabet: adenine (A), guanine (G), cytosine (C), and thymine (T).

In 2020, a group of over 50 companies and universities formed an international consortium called the DNA Data Storage Alliance. The alliance aims to “create and promote an interoperable storage ecosystem based on DNA as a storage medium.” As the field evolves, the group plans to be involved in the creation of specifications and standards to help end-users incorporate DNA-based storage into existing data-storage hierarchies.

Bioinformatics Market and Industry Trends

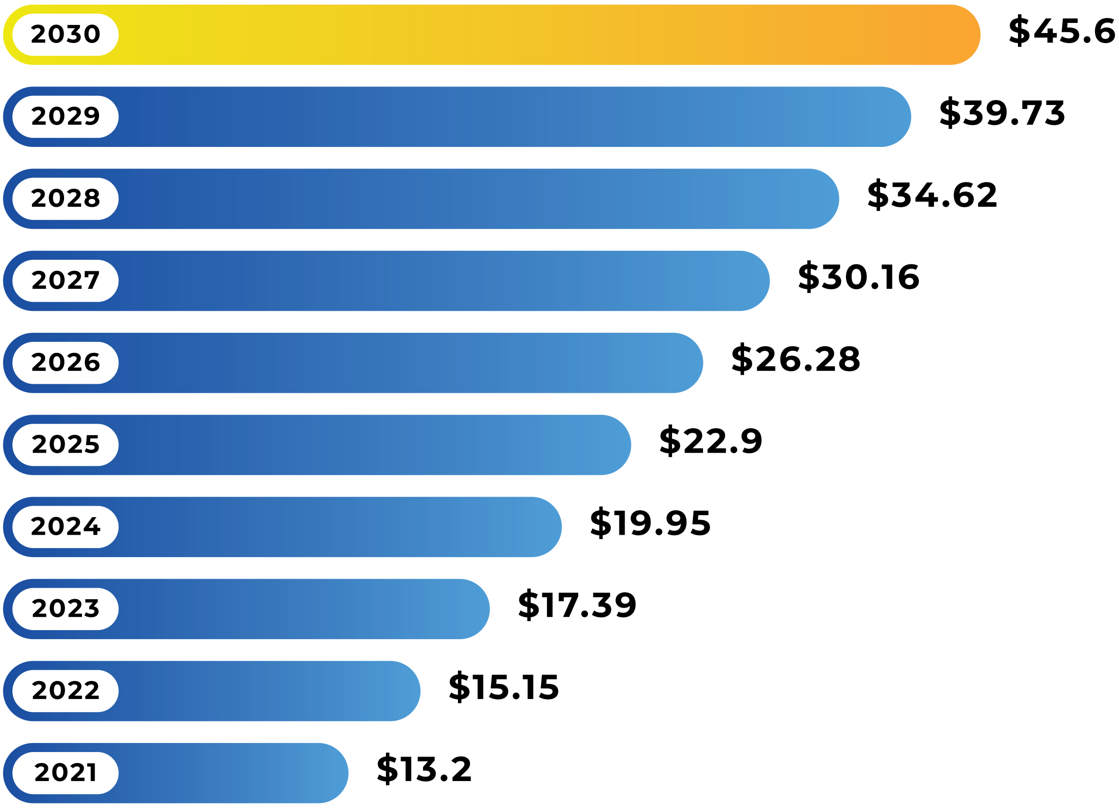
In May 2022, Precedence Research valued the 2021 global bioinformatics market at US$13.2 billion with the genomic segment of the field holding the largest market share. The genomic segment includes applications in pharmacogenomics, where bioinformatics is used to manage databases, and drug discovery, where bioinformatics is used for gene and protein sequencing.

The market analysis firm forecasts steady growth for the bioinformatics market, reaching about US$45.6 billion from 2022 to 2030—a compound annual growth rate of 16.3%. Per Precedence Research's analysts, North America's hold on the bioinformatics market is due to federal government support and funding for the field, as well as the presence of many established bioinformatics industries. Analysts also place Europe—with the United Kingdom and Germany at the head—as the fastest growing global market due to increased investments in the healthcare sector, as well as the introduction of artificial intelligence and the internet of things giving the euro market first-place speed. They named the United Kingdom and

Key Players in the Bioinformatics Industry

- Illumina Inc.
- Thermo Fischer Scientific Inc.
- QIAGEN N.V.
- Perkin Elmer Inc.
- Agilent Technologies Inc.
- DNAnexus
- Genentech
- Eurofins Scientific
- Waters Corporation
- NeoGenomics Laboratories
- Fios Genomics
- Source BioScience

Global Bioinformatics Market Forecast: 2021 – 2030
(USD billion)



Source: Precedence Research

German as the predicted top performers in the European market.

A recent report from market firm MordorIntelligence also forecasts sunny skies for the bioinformatics field. Their analysts predict the global market to grow at a compound annual growth rate of 10.9% from 2022 to 2027, with North America holding the largest share of the market and Europe the fastest growing market.

Driving growth in the global bioinformatics market appears to be the rising demand for DNA, RNA, and protein sequencing; rising initiatives from private and government organizations; advances in proteomics and genomics; and new research discoveries in molecular biology and drug discovery.

Growing concerns over data privacy and data security, as well as the high cost of bioinformatics tools and equipment, will potentially restrain growth in the global bioinformatics market.

Is a Career in Bioinformatics Right for You?

Most bioinformaticists have biology, physics, mathematics, and computer science backgrounds and typically work with scientists, clinicians, and engineers to solve biological problems and develop research tools.

Generally speaking, bioinformaticists either implement algorithms and develop tools for bioinformatics or are curators of data resources and data integration.

Is a Career in Bioinformatics Right for You?

A master's degree is usually required for a career in bioinformatics. Through formal education programs, individuals entering the bioinformatics field are expected to be familiar with concepts in biology, physics, mathematics, and computer programming and to develop in-depth knowledge in one of these sub-fields. The skills needed to become a bioinformaticist include, but are not limited to:

- A general understanding of biology
- In-depth knowledge of genomics and genetics
- In-depth knowledge of data mining and machine learning techniques, e.g., hierarchical clustering and decision trees
- Use of next-generation, high-throughput, and alignment sequencing tools, e.g., Bowtie, Partek, qPCR, and Ensemble
- Use of statistical software systems, e.g., SPSS and SAS
- Familiarity with programming languages, e.g., Perl, Python, Java, and Matlab
- Use of machine-learning tools and libraries, e.g., Millib and Scikit-Learn
- Use of database management tools for big-data databases, e.g., NoSQL, MonogoDB, TCGA, and Vertiga

Formal Education Programs in Bioinformatics

Bioinformatics Professional Associations

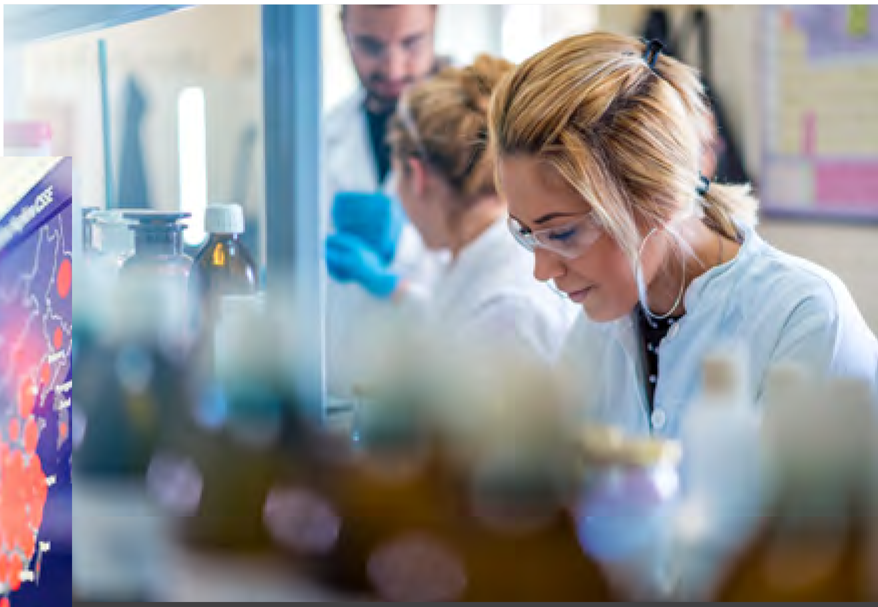
- The EMBL-European Bioinformatics Institute (EMBL-EBI): https://www.ebi.ac.uk
- The U.S. National Center for Biotechnology Information (NCBI): https://www.ncbi.nlm.nih.gov
- The National Institute of Genetics in Japan (NIG): https://www.nig.ac.jp/nig/
- International Society for Computational Biology: https://www.iscb.org

Bioinformatics Training Programs

The International Society for Computational Biology has compiled a database of formal undergraduate, master's, and doctorate degree programs. It is available at: www.iscb.org/iscb-degree-certificate-programs

Search our catalog of courses to advance your career in bioinformatics.

Careers in Bioinformatics



Once educated and trained, there are many opportunities to build a career in bioinformatics.

Bioinformaticists work in a variety of settings, including bioinformatics core facilities, biological and medical research laboratories, software development organizations, pharmaceutical and instrument development companies, and institutions that provide education and training.

Bioinformatics Professions

- Bioinformatics scientist: Develops software and databases for analyzing biological data.
- Research scientist: Conducts laboratory research in an academic environment.
- Biostatistician: Focuses on statistical design and analysis for scientific research applications.
- Microbiologist: Interprets anomalies and other patterns in research data to help identify ways bacteria and viruses impact human health.
- Bioinformatician: Manages large databases, develops data frameworks, and creates and modifies algorithms for classifying biological systems.
- Zoologist or wildlife biologist: Conducts in-depth analysis of animal population models, genetics, and the impact of changes in habitat.
- Molecular biologist: Uses bioinformatics skills to analyze large sets of genomic, proteomic, or imaging raw data.

Source: https://www.northeastern.edu/graduate/blog/top-bioinformatics-careers/

Examples of functions a bioinformaticist may perform include:

- Overseeing a laboratory information management system
- Designing strategies for DNA, RNA, and protein sequence analysis
- Developing algorithms to support next-generation DNA, RNA, and protein sequencing
- Conducting quantitative analysis of biological images
- Evaluating drug candidates for their value as targeted therapies
- Assisting the development of more efficient food production methods
- Developing systems for analyzing terrain using remote sensor data
- Creating data visualization for use in reports

Source: https://www.northeastern.edu/graduate/blog/top-bioinformatics-careers/

Tales From the Field

Two bioinformatics experts discuss their specific niche in the field and offer advice to scientists who are interested in pursuing a career in bioinformatics.

Manish Gupta

Manish Gupta is a professor at the Dhirubhai Ambani Institute of Information and Communication Technology in Gandhinagar, Gujarat, India (DA-IICT), where he leads the laboratory of natural information processing. Gupta is also a member of the DNA Storage Alliance—a network of people committed to creating and promoting an interoperable storage ecosystem based on DNA as a data storage medium. He holds undergraduate degrees in physics, chemistry, and mathematics, and a master's degree in mathematics from the University of Lucknow, India. He earned a Ph.D. in mathematics from the Indian Institute of Technology, Kanpur, India.

Gupta's research at DA-IICT is focused on data processing: deciphering the information processing principles in life (systems biology), building a computer out of biomolecules (synthetic biology), and developing methods for error correction in information processing.

“Bioinformatics is at the intersection of mathematics, computer science, and biology. The field was started as a development of computational tools for storing and processing ‘omics’ data,” says Gupta, “however, it has expanded in many directions to now include computational and systems biology, synthetic biology, and DNA computing.”

For those interested in entering the field, Gupta suggests “learning as much as possible in mathematics, computer science, and biology—but also chemistry, as it provides the hardware of biology.”

* * *

Parag Mallick

Parag Mallick is an associate professor at Stanford University, Calif., USA, and the founder and chief scientist at Nautilus Biotechnology. Mallick is an interdisciplinary scientist who has developed multi-scale approaches to accelerate the discovery of diagnostic and prognostic protein biomarkers. He holds an undergraduate degree in computer science from Washington University in St. Louis, Mo., USA, and a Ph.D. in chemistry and biochemistry from the University of California, Los Angeles (UCLA). Mallick helped start the bioinformatics program at UCLA and serves as an instructor.

For individuals interested in pursuing bioinformatics, Mallick recommends "having a language of familiarity to converse with those in each discipline—biology, math, and computer science."

However, "you also need to cultivate an expertise—biology, computer science, or mathematics—and then go deep into that area," adds Mallick. "You can be an individual innovator in that space, but as a bioinformaticist, you are really and most importantly a collaborator."

According to Mallick, those studying bioinformatics tend to gravitate toward one of four topical "buckets": tool developer, model developer, data analyst, and infrastructure builder.

"One of the hardest parts about being a bioinformaticist is that it is such a wide-ranging field. You always feel like you don't know enough," he shares. "You can be sitting in a meeting with a mathematician who is discussing really complex concepts you've never encountered. This is normal. Ultimately, as a bioinformaticist, you need to be comfortable with being uncomfortable."

IEEE
COMPUTER
SOCIETY

www.computer.org