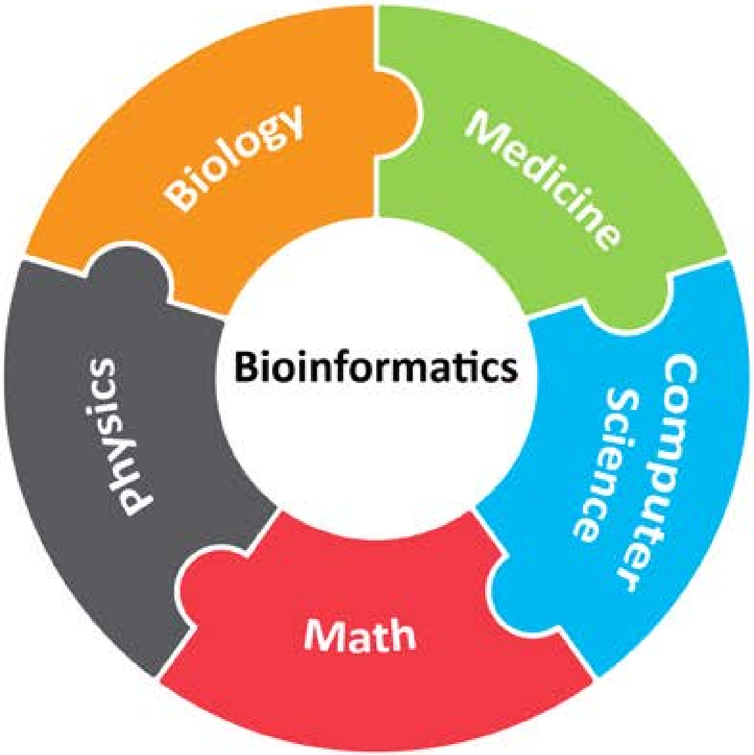
Bioinformatics – The Digital Future of Life Science Research & Development

Bioinformatics is one of the most rapidly evolving tools for accelerating life science research and biopharmaceutical development. As high throughput technologies become mainstream applications within the laboratory, the vast amount of data generated houses a wealth of information, providing you can unlock the door of understanding.

The term 'bioinformatics' was coined in the early 1970's by the Dutch research team of Paulien Hogeweg and Ben Hesper. It was first used to describe the study of information processes in biotic systems.[1] The roots of this emerging branch of biological science are evident as far back as the 1960's, as computational biologists focused on developing databases and algorithms to better interpret the rapidly accumulating data from protein biochemistry. Without the benefits of super computers or computer networks, these scientists laid important conceptual and technical foundations for bioinformatics today.[2]

This evolving branch of biological science is an interdisciplinary field that applies computer science to biology. New tools are emerging for the analysis of patterns that exist across DNA, RNA and protein sequences that assist in determining gene and protein functions, evolutionary biology, and 3-D modeling of proteins. For decades, bioinformatics has paralleled the development of high throughput DNA sequencing. One of the most prominent examples of the practical application of bioinformatics is the Human Genome Project (HGP), led by Dr. Francis Collins, Director of the National Health Institute (NIH). Originating in 1990, the HGP was chartered with providing a complete and accurate sequence of the 3 billion DNA base pairs that make up the human genome and to find all of the estimated 20,000 to 25,000 human genes.[3]

Fig. 1: The Multidisciplinary elements of Bioinformatics, an increasingly important tool of modern biological research



Through the collaborative efforts of 20 institutions in 6 countries, the task of mapping the human genome came to completion in 2003. The ambitious efforts of the HGP gave rise technological advances in DNA sequencing, and new tools for efficient data analysis and storage, shared with the global scientific community. This feat would not have been possible without the aid of bioinformatics and remains one of the main achievements in the field to date.

Shaping the Future of Healthcare Through Bioinformatics

The application of bioinformatics has since widely progressed into nearly every aspect of life science research and development, drug discovery, biopharmaceutical product development, and precision medicine. Researchers around the globe are using bioinformatics to identify genes and determine their functions to aid in the prevention, diagnosis and treatment of disease. This is achieved through the creation of databases and algorithms used to analyze enormous amounts of biological data. Mathematical models and dynamic simulations based upon pattern analysis of DNA or proteins are developed to help predict biological outcomes.

Examples of innovations in healthcare as a result of bioinformatics include:

Drug Discovery: The identification of drug targets and drug candidate screening process can be accelerated though the use of structure-based design, molecular modelling and simulation to develop safer and more effective drugs.

Gene Therapy: The primary goal of gene therapy is to replace defective genes with normal functioning genes for the treatment of genetic diseases that are non-responsive to traditional approaches in medicine. Bioinformatics enables the ability to identify the optimum gene target site for each individual, which can vastly improve the effectiveness of the treatment while reducing unwanted side effects.

Personalized & Preventive Medicine: Genetic testing to screen for susceptibility to genetic diseases or metabolic disorders are possible using bioinformatics tools to analyze genomic, proteomic or metabolomic data for the presence of gene mutations or potential disease biomarkers. The screening results are used to guide preventive measures or inform therapeutic treatment.

The Digital Building Blocks of Bioinformatics

Bioinformatics has opened the door to the elucidation of complex biological structures and processes, ultimately leading to the treatment or cure of genetic diseases. High throughput technologies, such as next-generation sequencing (NGS), have enabled the screening of large populations of biological samples. Biological databases also play a significant role in bioinformatics, allowing researchers to access a wide array of genomic sequences and other biologically relevant data from worldwide clinical and research efforts. Many databases and global resources exist that are open source tools available to the scientific community, such as the National Center for Biotechnology Information (NCBI) database, an important resource for bioinformatics tools. Other databases that play a key role in bioinformatics include the NIH GenBank DNA sequence database, and the cBio Cancer Genomics Portal which is funded in part by the National Cancer Institute (NCI) and provides visualization, analysis and download of large-scale cancer genomics data sets.[4,5]

However, the major challenges hindering the ability to effectively analyze information from these various sources is due to the inability to easily aggregate these disparate data sets. Multitudes of file formats exist across the global bioinformatics toolkit including .gb, .EMBL, .PDB, .MDL and .ABI for example. Data silos, differing file formats and non-standardized ontologies create integration challenges with the host of bioinformatic tools, making big data analytics an exceedingly difficult and inefficient task.

With the primary goal of bioinformatics being able to mine data from a wide variety of sources, analyzing and comparing it to a plethora of global databases for the presence of patterns within sequences, it is imperative to have a digitally connected laboratory and informatics environment. Digital transformation initiatives can be seen across nearly every segment of the scientific industry. Never before have the guiding principles underpinning this effort been so important as they relate to the field of bioinformatics. The key to unlocking the secrets that lie within these vast data lakes is the ability to effectively aggregate and mine scientific data and information from disparate sources through a seamless digital environment.

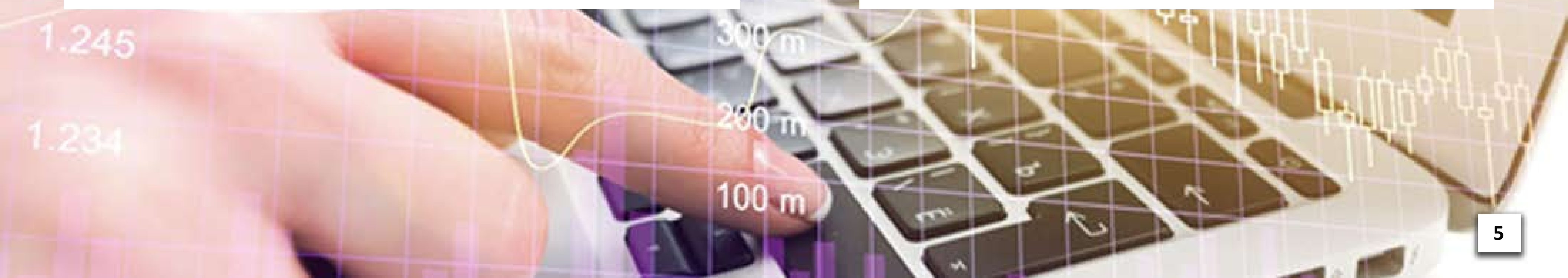
The Crucial Role of the Bioinformatician

As bioinformatics continues to play a vital role in life science research and development, having qualified personnel to conduct the needed multidisciplinary functions of the bioinformatician can be as equally challenging as the science itself. The diverse array of skill sets and expertise required to be effective in this role can include:

- Ph.D. in Bioinformatics, Genomics or a related discipline
- Post-doctoral experience in bioinformatics
- Expertise in Next Generation Sequencing (NGS) Data Analysis
- Working knowledge of statistical software for looking at gene expression patterns
- Ability to install, develop and maintain local version of publicly available bioinformatics tools
- Proficiency in software and algorithm development techniques
- Development of web-based genome browsers for analyzing and visualizing aggregate genome and exome data
- Experience programming in Perl, Python or equivalent scripting languages
- Development and deployment of applications on UNIX or Linux platforms in a high-performance computing (HPC) environment
- Communication and teaching skills to effectively train scientific staff on NGS and genomics analysis tools
- Technical writing competencies necessary for the preparation of user guides, manuals, and co-authoring of scientific manuscripts

While this only represents a subset of the technical and business acumen required of the bioinformatician, proficiency with informatics tools such as BLAST, JMP Genomics and BioPython, to name few, will often be essential to the success of most bioinformatics projects.

Utilizing a scientific staffing service with expertise in delivering uniquely qualified candidates for difficult to source positions will greatly improve the likelihood of finding the highest quality resources to fulfill the bioinformatics needs of your organization.

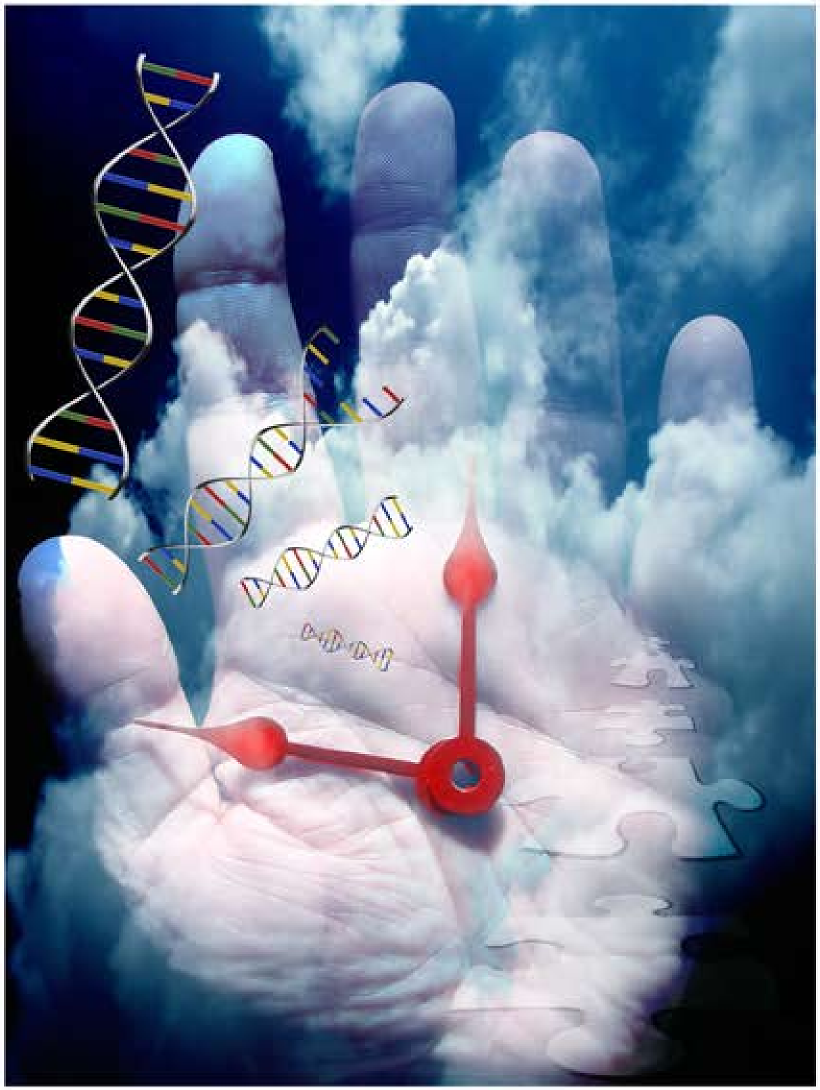


Summary

The world of bioinformatics is continuously evolving with new game-changing tools, techniques and informatics solutions rapidly becoming available, leading to the acceleration of scientific discovery and the advancement of medicine. As therapeutic treatment becomes more individualized, the future of healthcare is heavily reliant on our ability to understand the information that lies within the genetic code in order to provide the best patient care and prevent disease.

Bioinformatics greatly enhances our ability to analyze genetic sequences and interpret the functionality of genes, but in spite of the wealth of data that exists across the global scientific community, the process is significantly hindered by the siloed and disparate data warehoused in innumerable sources. Digital transformation holds the key to unlocking those secrets by providing the digital connectivity that enables ready access to data obtained from all available information sources using automated methodologies and advanced analytics to interpret aggregated data.

The ultimate piece of the puzzle is the crucial role of the bioinformatician in bringing all of the various components of the equation together to complete the picture. Their ability to aggregate, analyze, visualize and interpret data from a wide variety of sources provides scientists and clinicians the information needed to address biological questions and inform healthcare decisions.



References

[1]Hogeweg P (March 2011). Searls DB (ed.). "The roots of bioinformatics in theoretical biology". PLOS Computational Biology. 7 (3): e1002021. Bibcode:2011PLSCB...7E2021H. doi:10.1371/journal.pcbi.1002021. PMC 3068925. PMID 21483479.

[2]Hagen JB. The origins of bioinformatics. Nat Rev Genet. 2000 Dec;1(3):231-6. doi: 10.1038/35042090. PMID: 11252753.

[3]National Human Genome Institute, https://www.genome.gov/human-genome-project/results, accessed Sept.16, 2021.

[4]National Center for Biotechnology Information (NCBI), GenBank, https://www.ncbi.nlm.nih.gov/genbank/, accessed Sept.16, 2021.

[5]National Cancer Institute, Informatics Technology for Cancer Research (ITCR), cBioPortal for Cancer Genomics, https://itcr.cancer.gov/informatics-tools/cbioportal-cancer-genomics, accessed Sept.16, 2021.

About Astrix

For over 25 years, Astrix has been a market-leader in dedicated digital transformation & dedicated staffing services for science-based businesses. Through our proven laboratory informatics, digital quality & compliance, and scientific staffing services we deliver the highly specialized people, processes, and technology to fundamentally transform how science-based businesses operate. Astrix was founded by scientists to solve the unique challenges which science-based businesses face in the laboratory and beyond. We're dedicated to helping our clients speed & improve scientific outcomes to help people everywhere.