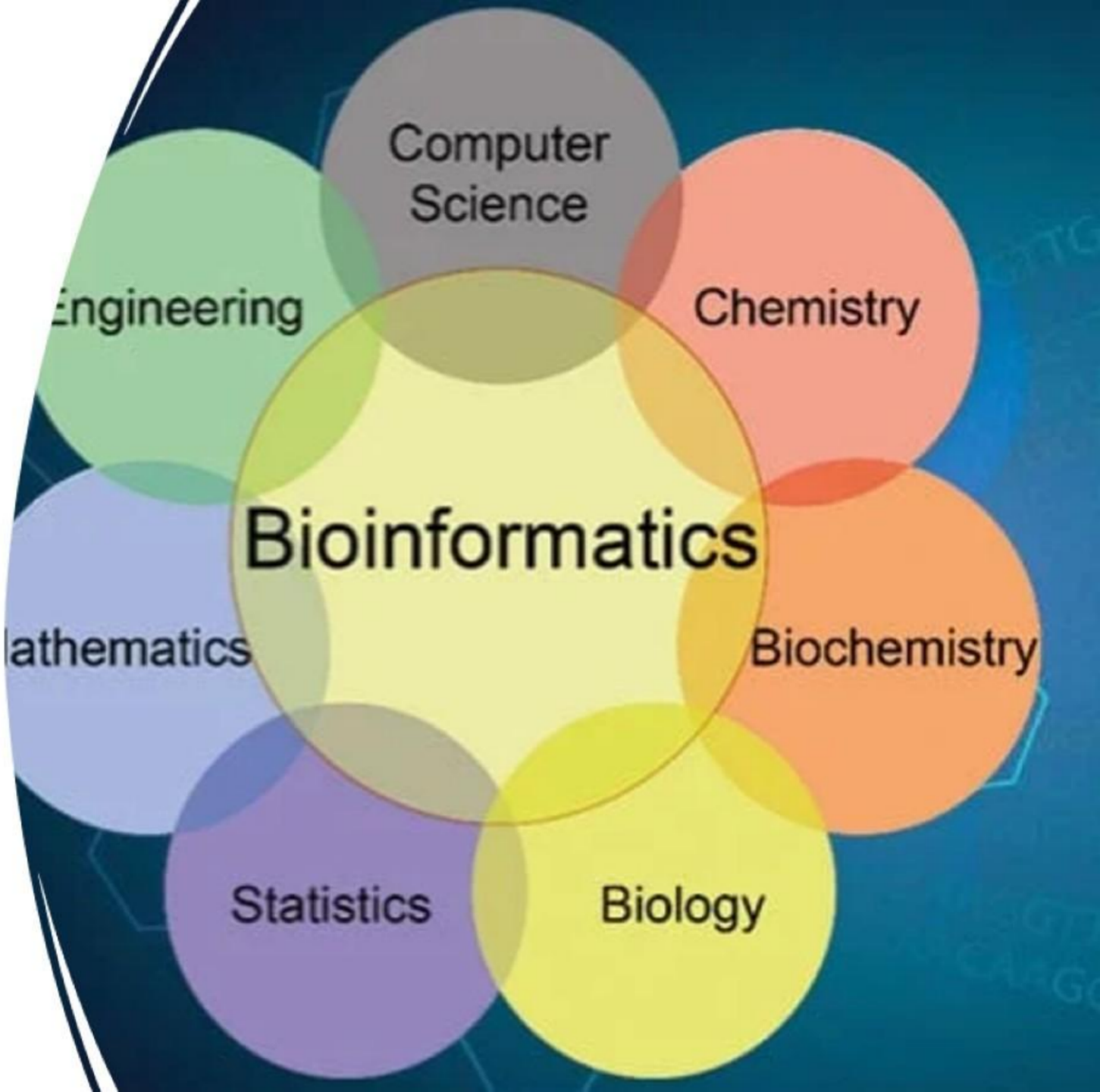
Bioinformatics Lec 1

Grade 3/ Biology Dept.

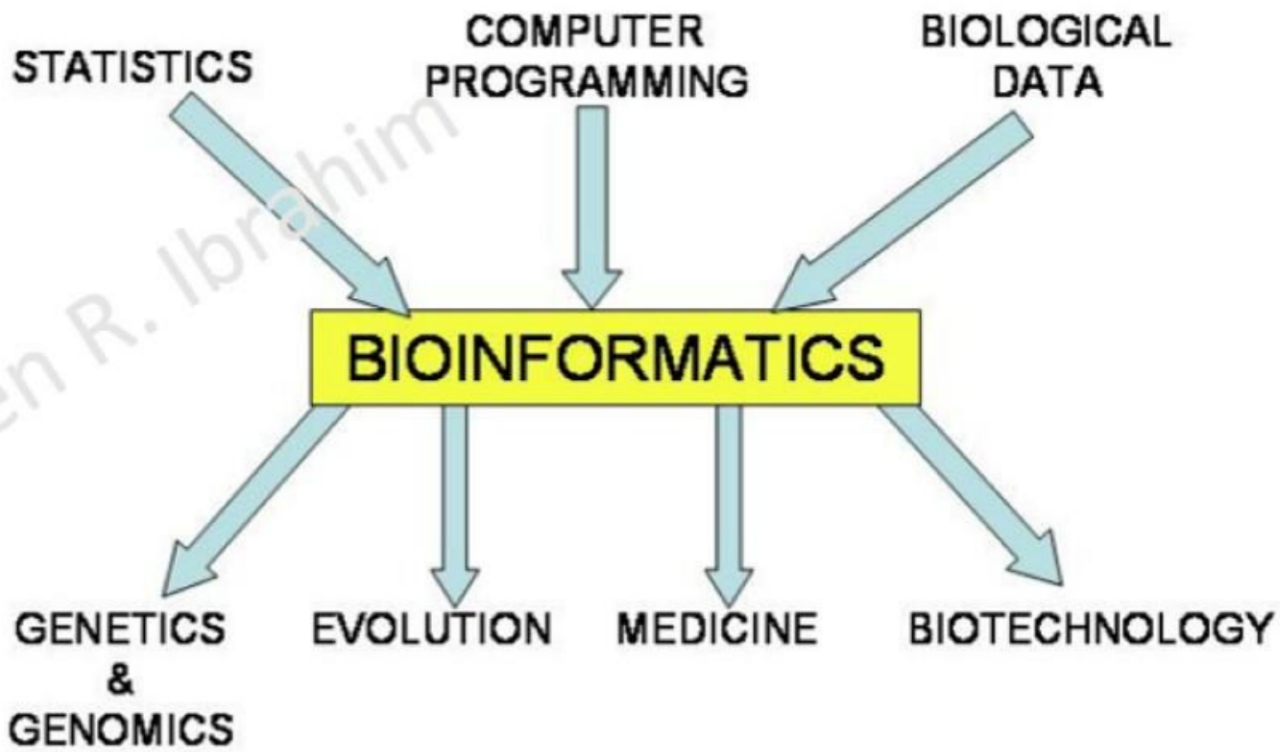
By

Dr Delveen R. Ibrahim



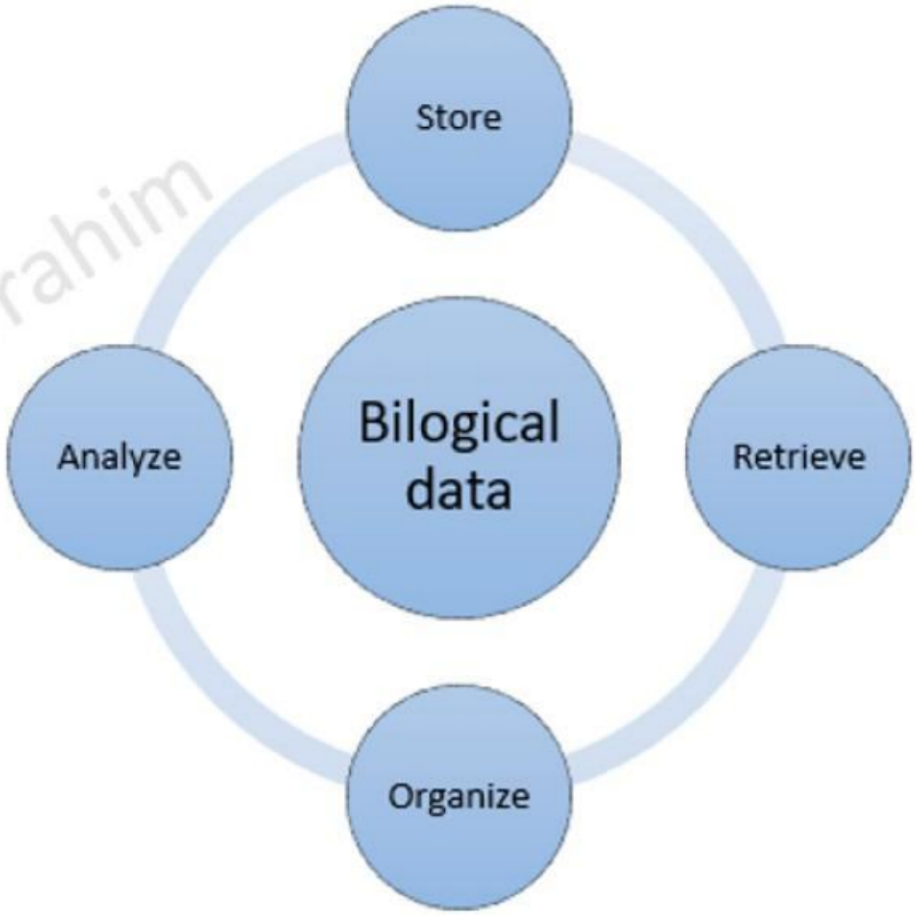
What is Bioinformatics??

- Bioinformatics is an interdisciplinary field that develops methods and software tools for understanding biological data.



The Building Blocks of Bioinformatics

- **Data in Biology**
  - DNA, RNA, and proteins as the language of life.
- **Role of Computers**
  - Storing and analyzing biological data.



Aims of bioinformatics

The field of bioinformatics has three main objectives:

1. **Data Organization**: Store, manage, and maintain biological data systematically in databases

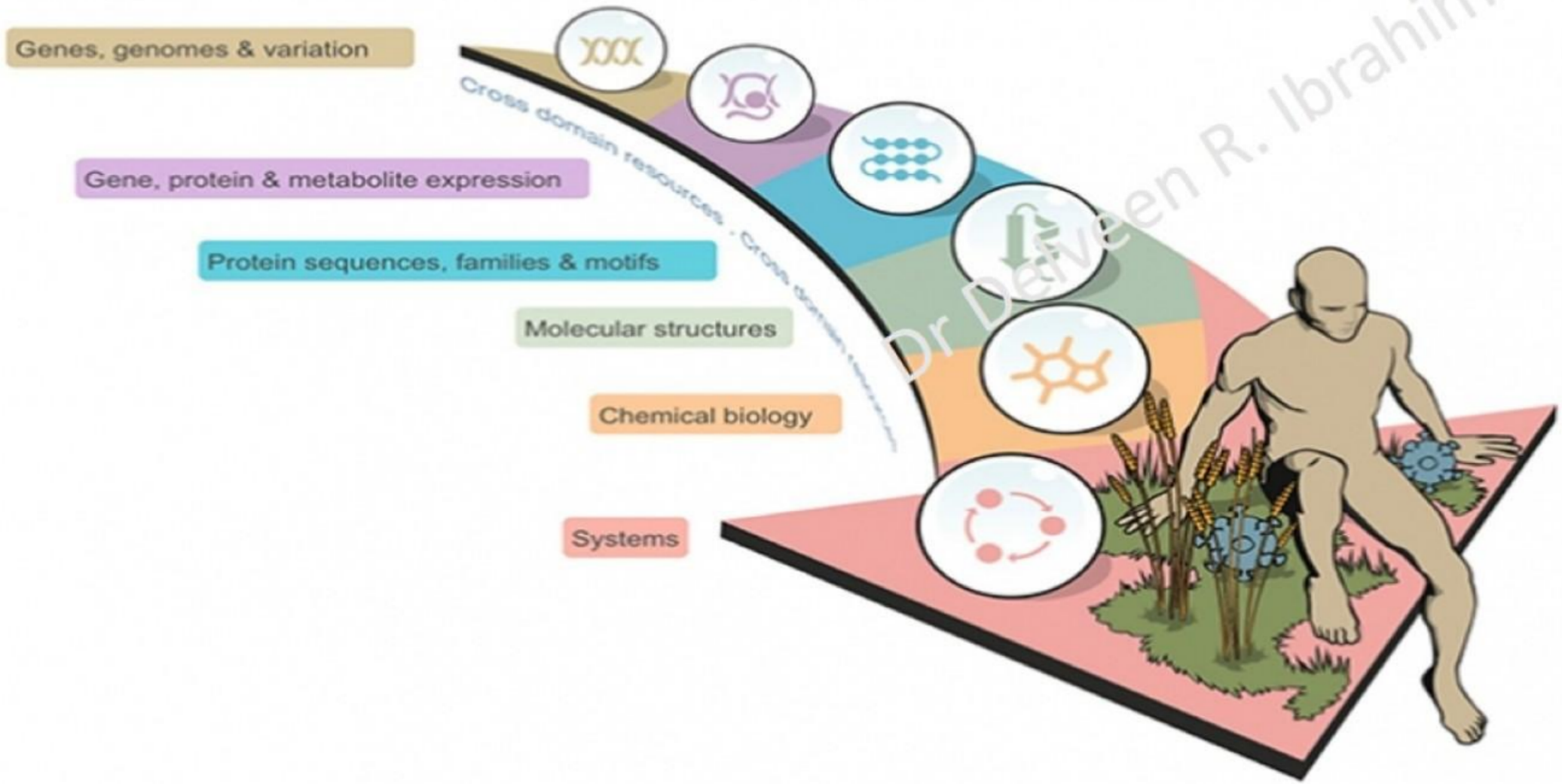
2. **Data Analysis**: Process and analyze biological data to extract meaningful insights.

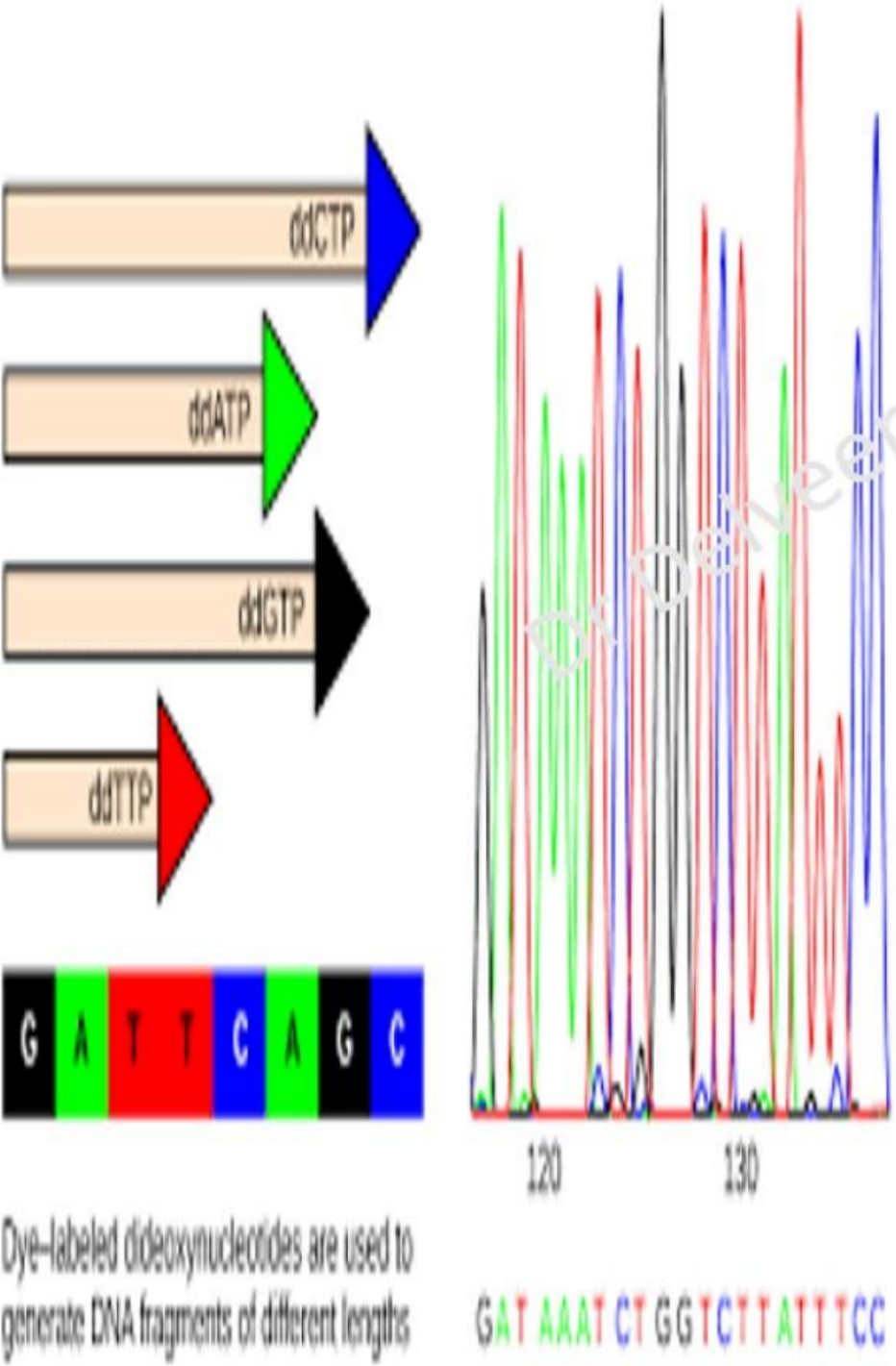
3. **Data Interpretation:** Understand the underlying biological significance of data

Types of biological data based on data type:

- **Genomic Data**: Involves the study of an organism's complete set of DNA, including genes and non-coding regions.
- **Transcriptomic Data**: Focuses on the study of RNA molecules, examining gene expression levels.
- **Proteomic Data**: Analyzes the structure and function of proteins within a biological system.
- **Metabolomic Data**: Explores the small molecules involved in metabolic pathways.
- **Biological literature data:** Organizing journals and articles

A broad overview of the different types of data that fall within the scope of bioinformatics





Dye-labeled dideoxynucleotides are used to generate DNA fragments of different lengths

Applications of bioinformatics:

- There are Many applications of Bioinformatics , here are some of the most important applications:

1. Health and Medicine

Application	Description	Examples
Genomics	Identifying genes responsible for diseases.	BRCA genes in breast cancer.
Drug Discovery and Development	Designing and simulating drug-target interactions.	COVID-19 antivirals.
Personalized Medicine	Tailoring treatments based on genetic makeup.	Pharmacogenomics studies.
Vaccine Development	Designing vaccines using pathogen genomes.	mRNA vaccines for COVID-19.

2. Agriculture

Application	Description	Examples
Crop Improvement	Identifying genes for traits like drought tolerance or pest resistance.	Bt cotton, Golden Rice.
Livestock Breeding	Enhancing breeding programs for healthier livestock.	High-yield dairy cows.
Food Safety	Detecting pathogens in food supplies.	Tracing foodborne outbreaks.

3. Environment

Application	Description	Examples
Microbial Analysis	Studying microbial composition of ecosystems.	Soil health improvement.
Bioremediation	Identifying organisms for pollutant degradation.	Cleaning oil spills.
Climate Change	Studying plant and animal adaptation to changing climates.	Drought-tolerant crops.

Others:

Evolution	Comparative Genomics	Understanding evolutionary relationships by comparing genomes.	Tracing organisms' ancestry.
Forensic Science	DNA Analysis	Identifying individuals from DNA at crime scenes.	DNA fingerprinting.
Education and Outreach	Teaching Tools	Online resources for teaching bioinformatics concepts.	NCBI tools, phylogenetic tree builders.

History of the bioinformatics

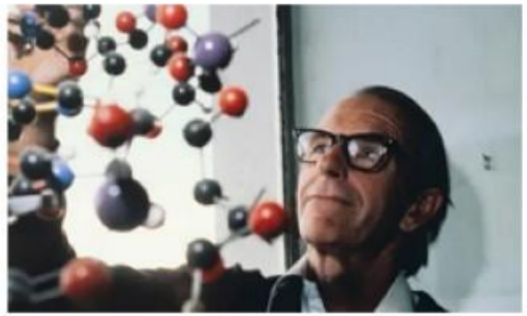
Term bioinformatics was invented by Paulien Hogeweg and Ben Hesper in 1970 as "the study of informatic processes in biotic system

While the first informatician is Margaret Dayhoff (1925–1983) , was an american physical chemist who pioneered the application of computational methods to the field of biochemistry

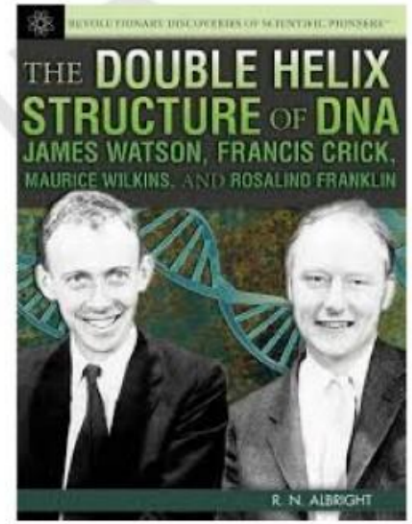
Paulien Hogeweg

Ben Hesper

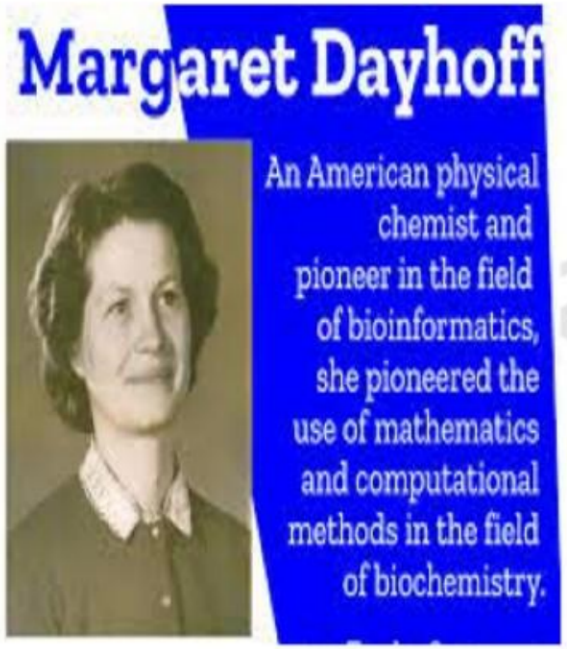
History of Bioinformatics



1952 Sanger sequenced bovine insulin



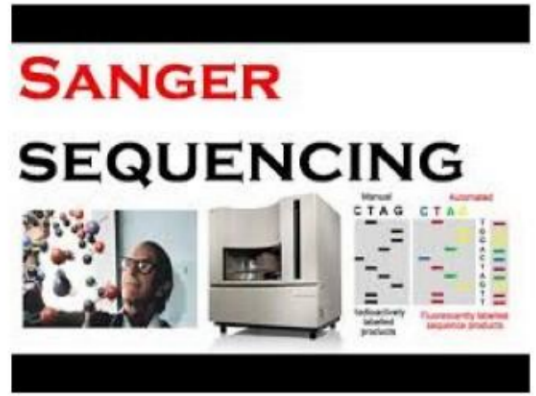
1953 Watson and Crick published DNA structure



1965 Margert Dayhoffs published protein sequence Atlas

1970 Paulien Hogeweg and Ben Hesper invented the term of Bioinformatics

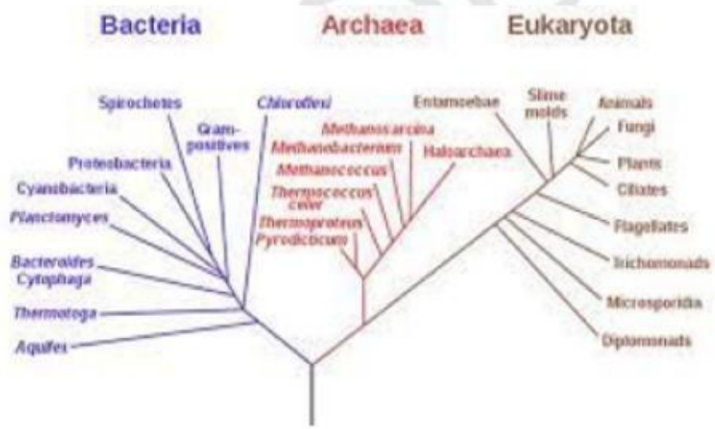
1977 DNA sequencing (Sanger) and software to analyze it (Staden)



History of Bioinformatics

1993 Sanger center for sequencing, UK

1995 *Hemophilus influnzae* genome Sequenced and *Methanococcus* genome sequenced , which confirmed the existence of third major branch of life on earth



In the 1990–2000s, use of the Internet, coupled with next-generation sequencing (NGS) , led to an exponential entry of data and a rapid increase of bioinformatics tools.

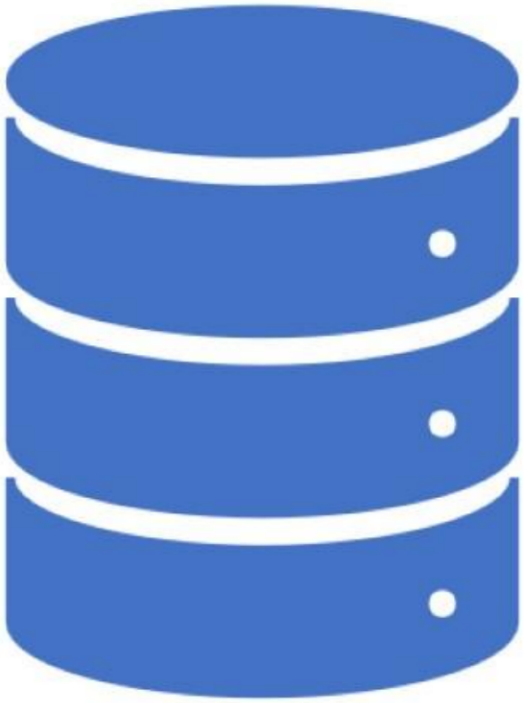
1990-2003 Human genome sequenced complete (13 years)

- **Modern Advances (2010s-Present)**
- **Advances in Sequencing Technologies**: Next-generation sequencing (NGS) enabled rapid, cost-effective genome sequencing.
- **Big Data and Machine Learning**: Bioinformatics involved artificial intelligence for analyzing massive datasets in genomics, proteomics, and systems biology.
- **AlphaFold (2020)**: A breakthrough in protein structure prediction, demonstrating the power of AI in solving biological problems.

Future of Bioinformatics

- **Career Opportunities:** Bioinformatician, genetic researcher, computational biologist.
  - How ?
  - By learning programming (Python, R) and biology together.

Practical Lec 1/ Biological Databases



Biological Databases

A brief history of biological databases

1965

Margaret Dayhoff created the Atlas of Protein Sequence and Structure

1980

The EMBL Data Library created a catalog of biological data

1982

GenBank became a public database.

2002

ArrayExpress released

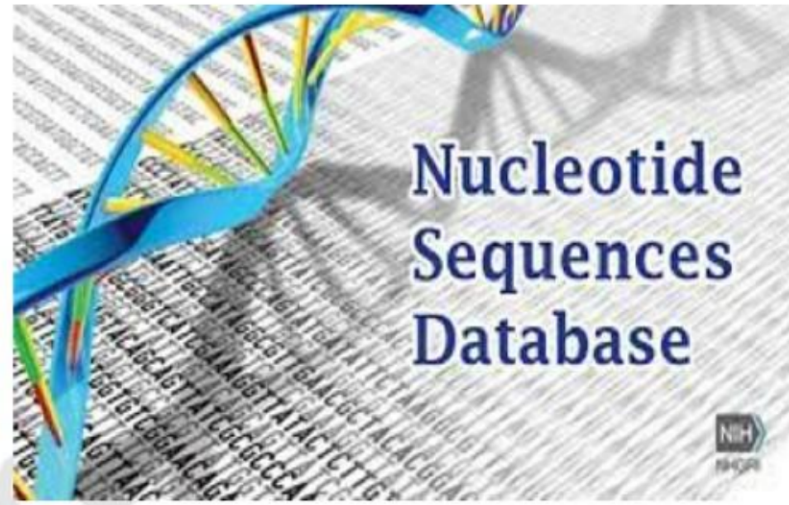
2005

The first high-throughput sequencing systems was used to sequence a full *E. coli* genome

PRESENT

Many directories for multi-omic data were created.
Let's explore some databases!

Biological databases



- Nucleotide and protein sequences databases can be classified based on accessibility to:
- **Primary databases**
- **Secondary databases**
- **Tertiary databases**

Primary databases

- It can also be called an archival database since it archives the experimental results submitted by the scientists. The primary database is populated with experimentally derived data like genome sequence, macromolecular structure, etc.
- It obtains unique data obtained from the laboratory and these data are made accessible to normal users without any change.
- The data are given accession numbers when they are entered into the database. The same data can later be retrieved using the accession number. Accession number identifies each data uniquely and it never changes.
- **Examples –**
- **Nucleic Acid Databases**: GenBank, EMBL and DDBJ
- **Protein Databases** are: PDB

Secondary databases

- The data stored in these types of databases are the analyzed result of the primary database.
- The data here are highly curated (processing the data before it is presented in the database). A secondary database is better and contains more valuable knowledge compared to the primary database.
- **Examples –**
- **DNA databases**: Refseq (**comprehensive, integrated, well-annotated set of sequences, including genomic DNA)**
- SNP- disease databases
- **Protein Databases**:
- InterPro (protein families, motifs, and domains)
- UniProt/KB (sequence and functional information on proteins)

Based on Accessibility	Primary Databases	Contain raw experimental data submitted by researchers.	GenBank, PDB, GEO
	Secondary Databases	Contain curated and processed data derived from primary databases.	UniProt, Pfam
	Tertiary Databases	Integrate data from multiple primary and secondary	KEGG, Ensembl

Abbreviation

- NCBI: National Center for Biotechnology information
- EMBL: European Molecular Biology Library
- DDBJ: DNA Databank of Japan
- UniProt/KB: Universal Protein Knowledgebase
- PIR: Protein Information Resources
- PDB: Protein Data Bank
- NRDB: Non-Redundant Databases
- PubMed: Public/Publisher MEDLINE
- NLM: National library of medicine
- PMC: PubMed Central
- BLAST: The Basic Local Alignment Search Tool

How Journals and Articles Are Organized in Biological Databases

- Organizing journals and articles is an integral part of many biological databases, as they help researchers stay updated and access relevant literature for their studies. This is commonly achieved through **literature databases** or **integrated resources** within bioinformatics systems

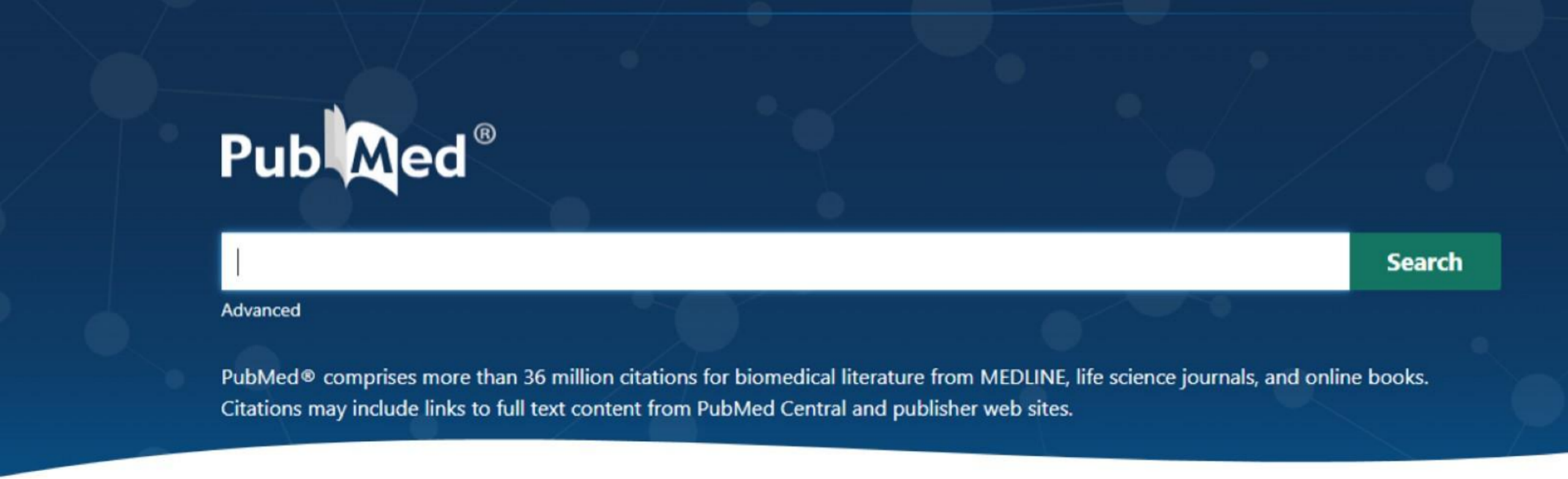
- **Specialized Literature Databases**
- Databases like **PubMed**, **Europe PMC**, and **Medline** specialize in indexing biological and medical research articles.
- They organize articles by:
  - Titles, authors, and abstracts.
  - Keywords and MeSH (Medical Subject Headings) terms for easy retrieval.
  - Citations and cross-references to other articles.

Examples of Biological Databases Integrating Literature

- **PubMed**: Focuses on biomedical and life sciences literature.
- **Europe PMC**: Provides access to full-text articles and links to datasets.
- **NCBI Gene**: Links genes to PubMed articles and related research.
- **UniProt**: Links protein entries to peer-reviewed publications explaining their properties.

Benefits of Organizing Articles in Biological Databases:

- Facilitates literature review for researchers.
- Enhances the reliability of biological data by linking it to peer-reviewed evidence.
- Saves time by combining experimental data and relevant studies in one place.



How to use PubMed and PMC

What is the difference between PMC and PubMed?

Dr Delveen Ramdan Ibrahim

National Library of Medicine
National Center for Biotechnology Information

delveen.ibrahim@uod...

PubMed®

Search

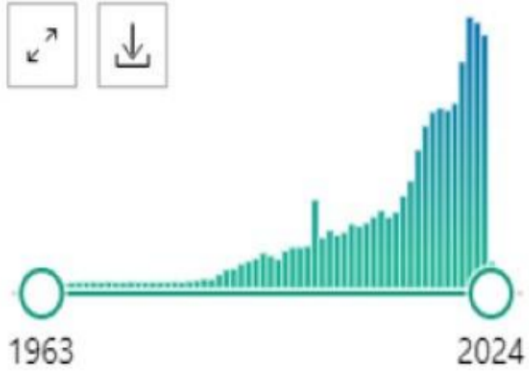
Advanced

PubMed® comprises more than 36 million citations for biomedical literature from MEDLINE, life science journals, and online books. Citations may include links to full text content from PubMed Central and publisher web sites.

MY NCBI FILTERS

25,843 results

RESULTS BY YEAR



Filters applied: Review. Clear all

1

Cite

Share

Compare and contrast: **pediatric cancer** versus adult malignancies.

Kattner P, Strobel H, Khoshnevis N, Grunert M, Bartholomae S, Pruss M, Fitzel R, Halatsch ME, Schilberg K, Siegelin MD, Peraud A, Karpel-Massler G, Westhoff MA, Debatin KM.

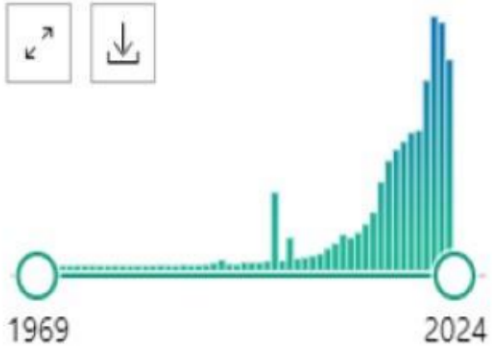
Cancer Metastasis Rev. 2019 Dec;38(4):673-682. doi: 10.1007/s10555-019-09836-y.

PMID: 31832830 Review.

MY NCBI FILTERS

10,311 results

RESULTS BY YEAR



Filters applied: Free full text, Review. Clear all

1 Recent Advances in **Pediatric Cancer** Research.

McEachron TA, Helman LJ.

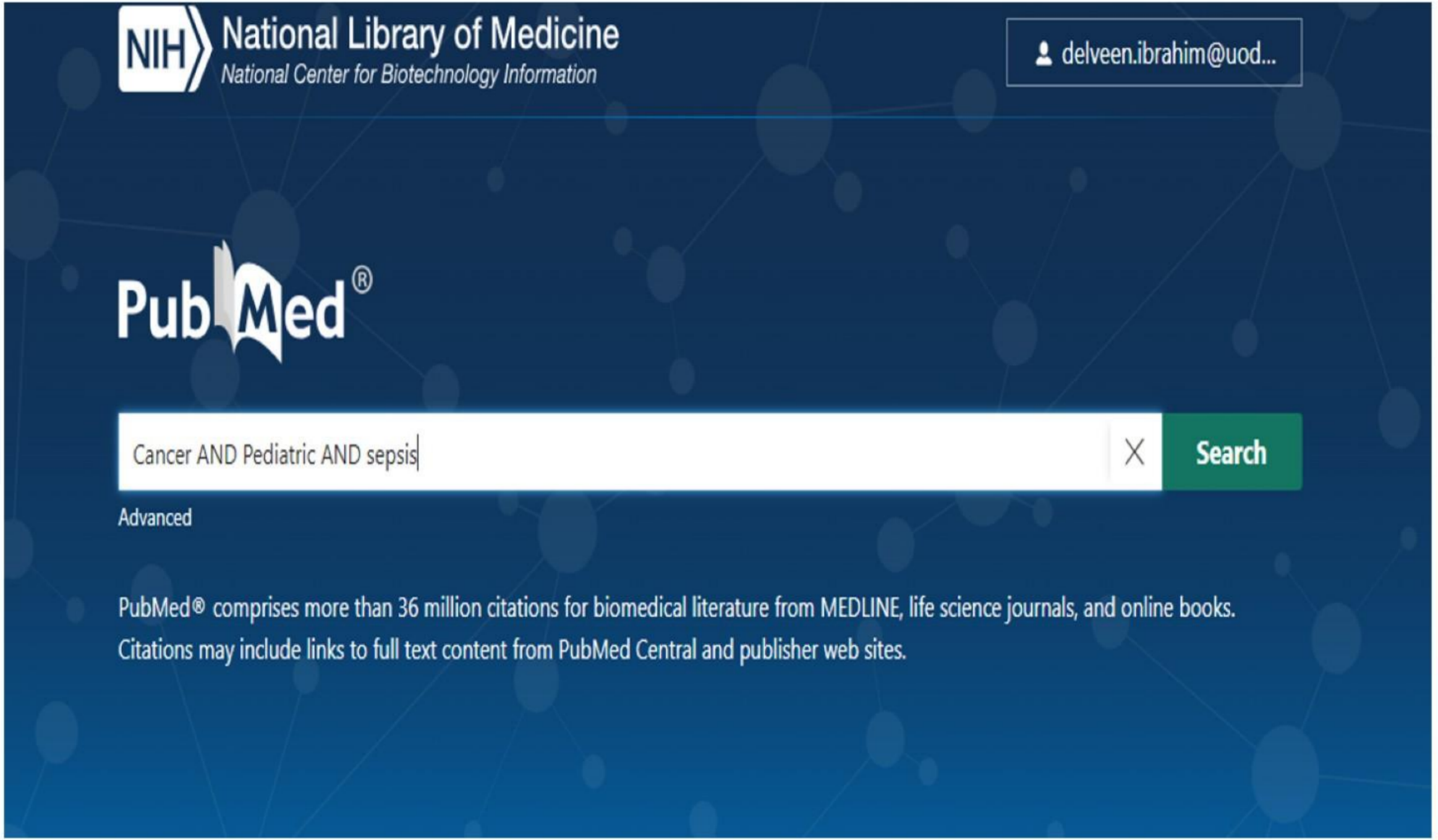
Cite

Cancer Res. 2021 Dec 1;81(23):5783-5799. doi: 10.1158/0008-5472.CAN-21-1191. Epub 2021 Sep 24.

PMID: 34561271 Free PMC article. Review.

Share

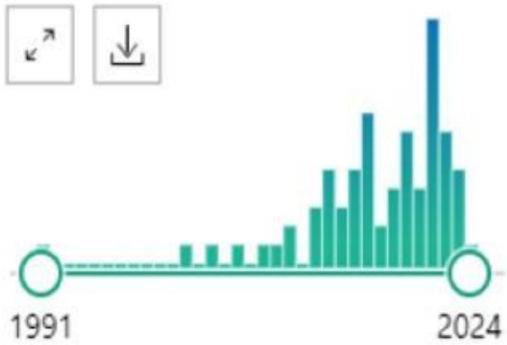
Over the past few years, the field of **pediatric cancer** has experienced a shift in momentum, and this has led to new and exciting findings that have relevance beyond **pediatric** malignancies. Here we present


National Library of Medicine
National Center for Biotechnology Information
delveen.ibrahim@uod...
PubMed®
Cancer AND Pediatric AND sepsis
Search
Advanced
PubMed® comprises more than 36 million citations for biomedical literature from MEDLINE, life science journals, and online books.
Citations may include links to full text content from PubMed Central and publisher web sites.

MY NCBI FILTERS

70 results

RESULTS BY YEAR



Filters applied: Free full text, Review. Clear all

1

Cite

Share

Beyond Hemostasis: Platelet Innate Immune Interactions and Thromboinflammation.

Mandel J, Casari M, Stepanyan M, Martyanov A, Deppermann C.

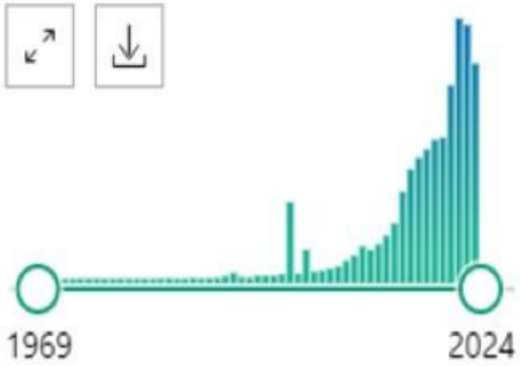
Int J Mol Sci. 2022 Mar 31;23(7):3868. doi: 10.3390/ijms23073868.

PMID: 35409226 Free PMC article. Review.

There is accumulating evidence that platelets play roles beyond their traditional functions in thrombosis

MY NCBI FILTERS

RESULTS BY YEAR



TEXT AVAILABILITY

- [ ] Abstract
- [x] Free full text
- [ ] Full text

ARTICLE ATTRIBUTE

- [ ] Associated data

ARTICLE TYPE

10,311 results

Page 1 of 1,032

Filters applied: Free full text, Review. Clear all

- [ ] Recent Advances in **Pediatric Cancer** Research.

1 McEachron TA, Helman LJ.

Cite Cancer Res. 2021 Dec 1;81(23):5783-5799. doi: 10.1158/0008-5472.CAN-21-1191. Epub 2021 Sep 24.

PMID: 34561271 Free PMC article Review.

ncer has experienced a shift in momentum, and this has

nce beyond **pediatric** malignancies. Here we present

CITE

McEachron, T. A., & Helman, L. J. (2021). Recent Advances in Pediatric Cancer Research. *Cancer research*, *81*(23), 5783–5799. https://doi.org/10.1158/0008-5472.CAN-21-1191

Copy Download .nbib Format: APA

immunogenomics.

10.1016/j.gde.2022.101918. Epub 2022 May 23.

cers established the importance of multiple platform-

ain a complete molecular landscape of these **cancers**

nostic eviden ...

National Library of Medicine
National Center for Biotechnology Information

PMC PubMed Central®

Search PMC Full-Text Archive

Search in PMC

Advanced | Journal List

PubMed Central® (PMC) is a free full-text archive of biomedical and life sciences journal literature at the U.S. National Institutes of Health's National Library of Medicine (NIH/NLM)

About PMC

Discover a digital archive of scholarly articles, spanning centuries of scientific research.

User Guide

Learn how to find and read articles of interest to you.

Collections

Browse the PMC Journal List or learn about some of PMC's unique collections.

Article attributes
Associated Data
Author manuscripts
Digitized back issues
MEDLINE journals
Open access
Preprints
Retracted

Text availability
Include embargoed articles

Publication date
1 year
5 years
10 years
Custom range...

Display Settings: Summary, 20 per page, Sorted by Default order

Send to:

You searched **8+ million full text articles**

Try this search in **34+ million citations and abstracts**

PMC Full-Text Search Results

Items: 1 to 20 of 136619

<< First < Prev Page 1 of 6831 Next > Last >>

SIOP ABSTRACTS

1. **Pediatr Blood Cancer**. 2018 Nov; 65(Suppl 2): e27455. Published online 2018 Sep 21. doi: 10.1002/pbc.27455
PMCID: PMC7167781
Article PDF–7.1M

Filter your results:

All (136619)
PubMed links (131424)
Open Access (74329)
Embargoed (0)

Manage Filters

Find related data

Database: Select

Find items

Article attributes
Associated Data
Author manuscripts
Digitized back issues
MEDLINE journals
Open access
Preprints
Retracted

Text availability
Include embargoed articles

Publication date
1 year
5 years
10 years
Custom range...

Research Funder
NIH
AHRQ
ACL
ASPR

You searched **8+ million full text articles**

Try this search in **34+ million citations and abstracts**

PMC Full-Text Search Results

Items: 1 to 20 of 11975

1. Surviving **sepsis** campaign international guidelines for the management of septic shock and **sepsis**-associated organ dysfunction in children

Scott L. Weiss, Mark J. Peters, Waleed Alhazzani, Michael S. D. Agus, Heidi R. Flori, David P. Inwald, Simon Nadel, Luregn J. Schlapbach, Robert C. Tasker, Andrew C. Argent, Joe Brierley, Joseph Carcillo, Enitan D. Carrol, Christopher L. Carroll, Ira M. Cheifetz, Karen Choong, Jeffry J. Cies, Andrea T. Cruz, Daniele De Luca, Akash Deep, Saul N. Faust, Claudio Flauzino De Oliveira, Mark W. Hall, Paul Ishimine, Etienne Javouhey, Koen F. M. Joosten, Poonam Joshi, Oliver Karam, Martin C. J. Kneyber, Joris Lemson, Graeme MacLaren, Nilesh M. Mehta, Morten Hylander Møller, Christopher J. L. Newth, Trung C. Nguyen, Akira Nishisaki, Mark E. Nunnally, Margaret M. Parker, Raina M. Paul,

Filter your results:

All (11975)

PubMed links (11361)

Open Access (7260)

Embargoed (0)

Manage Filters

Find related data

Database: Select

Find items

Search details

("neoplasms"[MeSH Terms] OR "neoplasms"[All Fields] OR "cancer" [All Fields]) AND ("paediatrics"[All Fields] OR "pediatrics"[MeSH Terms]

Literature Search and Retrieval

- **Task**: Use PubMed or PMC to find articles related to a specific biological topic/ chose your project title.
- **Example**:
  - Search for papers on "CRISPR-Cas9 applications in gene therapy."
  - Refine results using filters like publication date, author, journal, or study type.