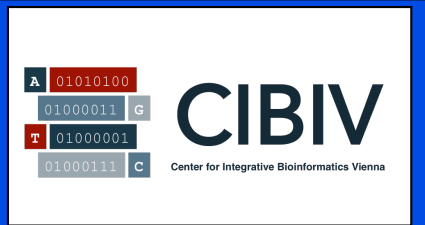


Bioinformatics: An Introduction

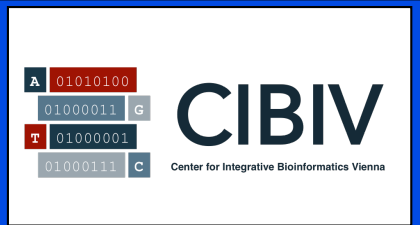


Heiko Schmidt

Center for Integrative Bioinformatics Vienna (CIBIV)

<http://www.cibiv.at>

What is Bioinformatics?

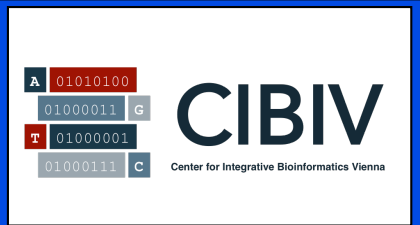


Bioinformatics is the field of science in which **biology**, **computer science**, and **information technology** merge to form a single discipline.

The ultimate goal of the field is to enable the **discovery of new biological insights** as well as to create a **global perspective** from which unifying principles in biology can be discerned.

Source: “NCBI homepage”

History of Bioinformatics

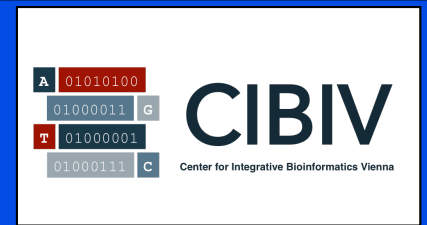


At beginning: ("genomic revolution"), the creation and maintenance of a database to **store biological information** (DNA, protein etc.) Computer Science challenges: design issues, complex interfaces (Access, revision, submission).

Goal: information must be combined to form a **comprehensive picture** of normal cellular activities.

New Tasks: **analysis and interpretation** of various types of data, including nucleotide and amino acid sequences, protein domains, and protein structures.

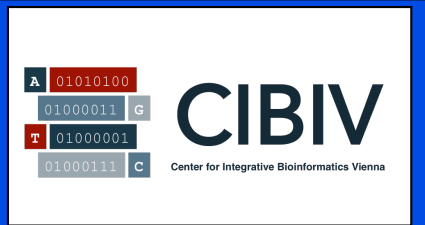
History of Bioinformatics



New (sub) discipline: **Computational Biology**:
analyzing and interpreting data:

- **Development and implementation** of tools for efficient access to, and use and management of, various types of information
- Development of **new algorithms** and **statistics** to assess relationships among members of large data sets,
 - to locate a gene within a (genomic) sequence
 - to predict protein structure and function
 - to cluster protein sequences into families of related sequences

Goals of today's presentation

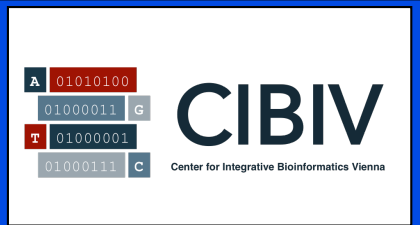


Brief overview of science in bioinformatics

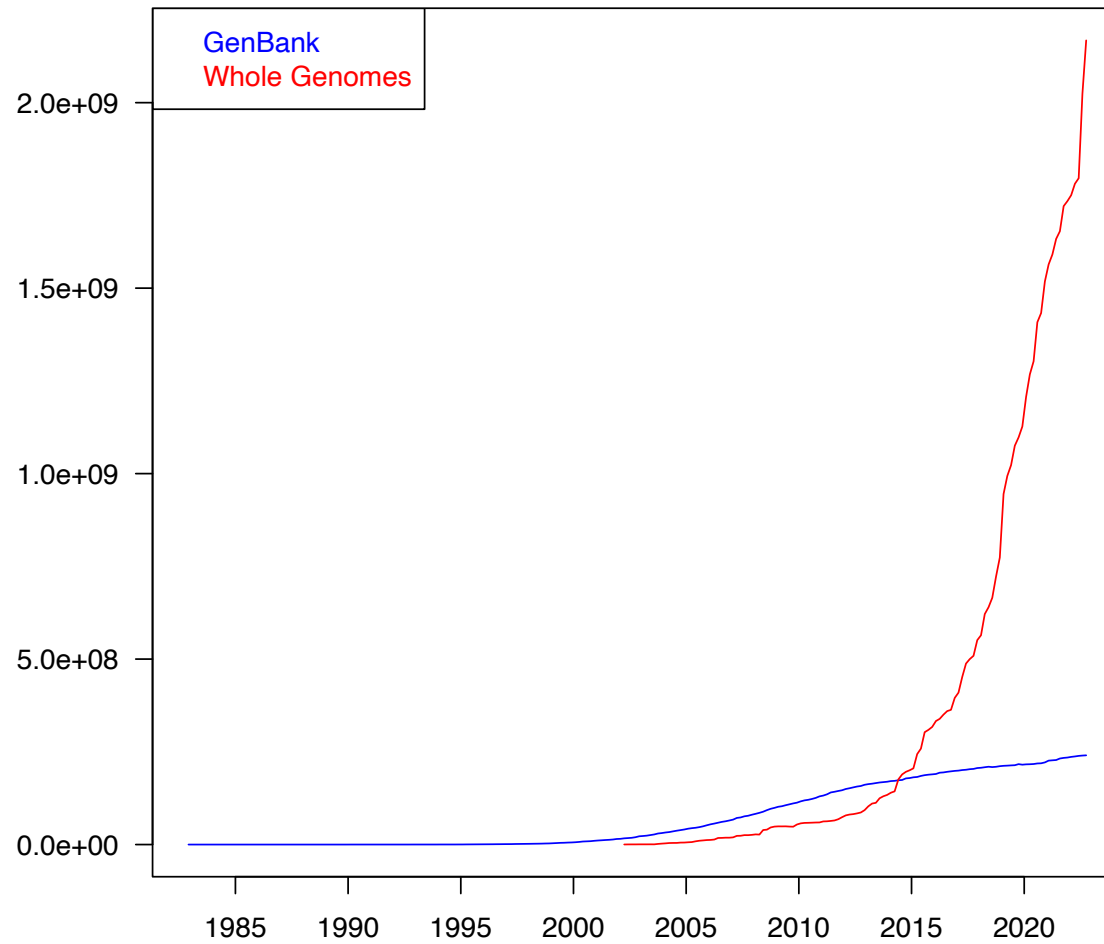
Databases and tools

Basics of sequence search and analysis

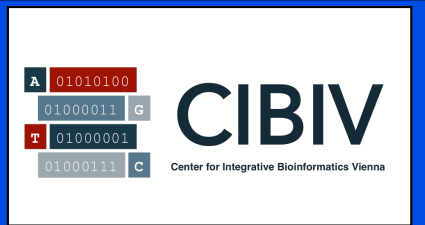
Why bioinformatics?



GenBank growth (number of sequences)

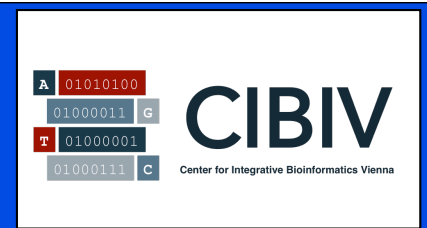


Fields in Bioinformatics



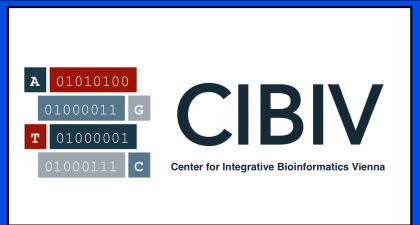
- Bioinformatics
- Molecular Genetics
- High-throughput Sequencing Analysis
- Genome Mapping
- SNP analyses
- Transcriptome Analysis (mRNAs)
- Metagenomics
- Pharmacogenomics
- Phylogenetics
- Phylogenomics
- Molecular Modeling

Types of Databases and Tools



- Nucleotide Databases
- Protein Databases
- Genome Databases
- Genome-Specific Resources
- Structure Databases
- Literature Databases
- Tools for Sequence Analysis
- Tools for Data Mining
- Tools for 3-D Structure Display and Similarity Searches
- Genomic Maps
- FTP Download Sites

Nucleotide Databases



INSDC: The 3 major public DNA databases (Intl. Nucleotide Sequence Database Collaboration)

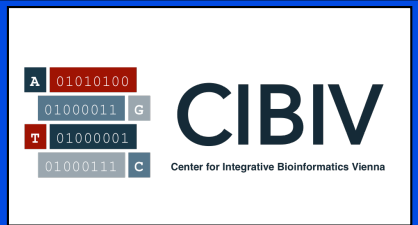


**European
Nucleotide
Archive**
housed
at EBI
**European
Bioinformatics
Institute**
(Hinxton, UK)

Housed
at NCBI
**National
Center for
Biotechnology
Information**
(USA)

**DNA
DataBank
of Japan**
housed
at NIG
**National
Institute of
Genetics**
(Japan)

Nucleotide Databases



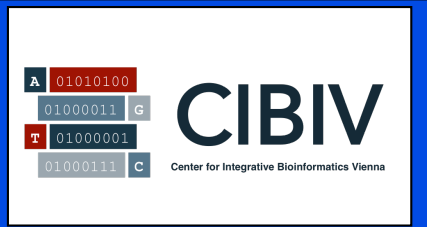
| | |
|-----------------------------|---|
| GenBank | An annotated collection of all publicly available nucleotide and amino acid sequences |
| Sequence Read Archive (SRA) | sequencing data from high-throughput sequencing platforms like Illumina, Roche 454, PacBio |
| Genome | sequence and map data from whole genomes of organisms (11.5k Archaea, 1.5mio Bacteria, 37.8k Eukaryotes, 1050 Viruses), complete and in progress |
| HomoloGene | A gene homology database that compares nucleotide sequences between pairs of organisms to identify putative orthologs |
| Taxonomy | names and phylogenetic lineages of more than 540,000 organisms |

Nucleotide Databases



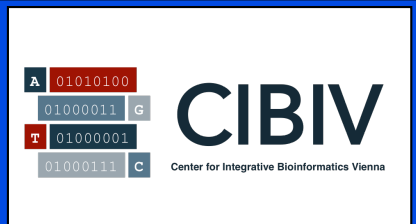
| | |
|---------|---|
| dbSNP | A central repository for both single-base nucleotide substitutions and short deletion and insertion polymorphisms |
| RefSeq | non-redundant reference sequences standards (including genomic DNA, mRNAs, and proteins for known genes) for genome annotation, gene identification, and comparative analyses |
| Gene | Gene supplies gene-specific information and may integrate nomenclature, Reference Sequences (RefSeqs), maps, pathways, variations, phenotypes, and links to genome-, phenotype-, and locus-specific resources worldwide. |
| ClinVar | archive of reports of clinically relevant human genetic variants and their relationships to phenotypes, with supporting evidence |

NCBI



National Center for Biotechnology Information (NCBI)

www.ncbi.nlm.nih.gov



NCBI Home

Resource List (A-Z)

All Resources

Chemicals & Bioassays

Data & Software

DNA & RNA

Domains & Structures

Genes & Expression

Genetics & Medicine

Genomes & Maps

Homology

Literature

Proteins

Sequence Analysis

Taxonomy

Training & Tutorials

Variation

Welcome to NCBI

The National Center for Biotechnology Information advances science and health by providing access to biomedical and genomic information.

[About the NCBI](#) | [Mission](#) | [Organization](#) | [NCBI News & Blog](#)

Submit

Deposit data or manuscripts into NCBI databases



Download

Transfer NCBI data to your computer



Learn

Find help documents, attend a class or watch a tutorial



Develop

Use NCBI APIs and code libraries to build applications



Analyze

Identify an NCBI tool for your data analysis task



Research

Explore NCBI research and collaborative projects



Popular Resources

[PubMed](#)

[Bookshelf](#)

[PubMed Central](#)

[BLAST](#)

[Nucleotide](#)

[Genome](#)

[SNP](#)

[Gene](#)

[Protein](#)

[PubChem](#)

NCBI News & Blog

Posters, Presentations, and a Booth: NCBI at ASM Microbe 2019

13 Jun 2019

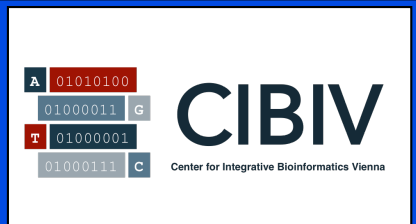
From June 20-24, 2019, NCBI staff will participate in ASM Microbe 2019. Not

Presentation on NCBI's genome browser at Rocky Mountain Genomics Hackcon

12 Jun 2019

On June 18, 2019, NCBI's Sanjida

NCBI PubMed



NIH National Library of Medicine
National Center for Biotechnology Information

Log in

PubMed.gov

Advanced

PubMed® comprises more than 34 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.



Learn

- [About PubMed](#)
- [FAQs & User Guide](#)
- [Finding Full Text](#)



Find

- [Advanced Search](#)
- [Clinical Queries](#)
- [Single Citation Matcher](#)



Download

- [E-utilities API](#)
- [FTP](#)
- [Batch Citation Matcher](#)



Explore

- [MeSH Database](#)
- [Journals](#)

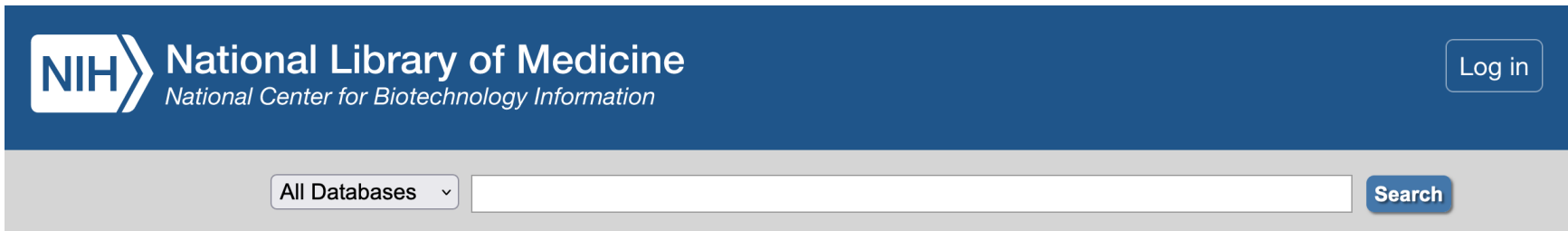
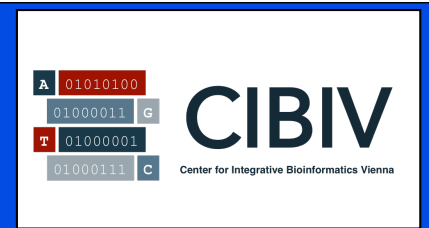


PubMed is...

National Library of Medicine's search service

- in citations and abstracts for biomedical literature from MEDLINE, life science journals, and online books.
- covers a total of **over 34 million entries** (reference and abstracts)
- links to participating online journals

The NCBI System

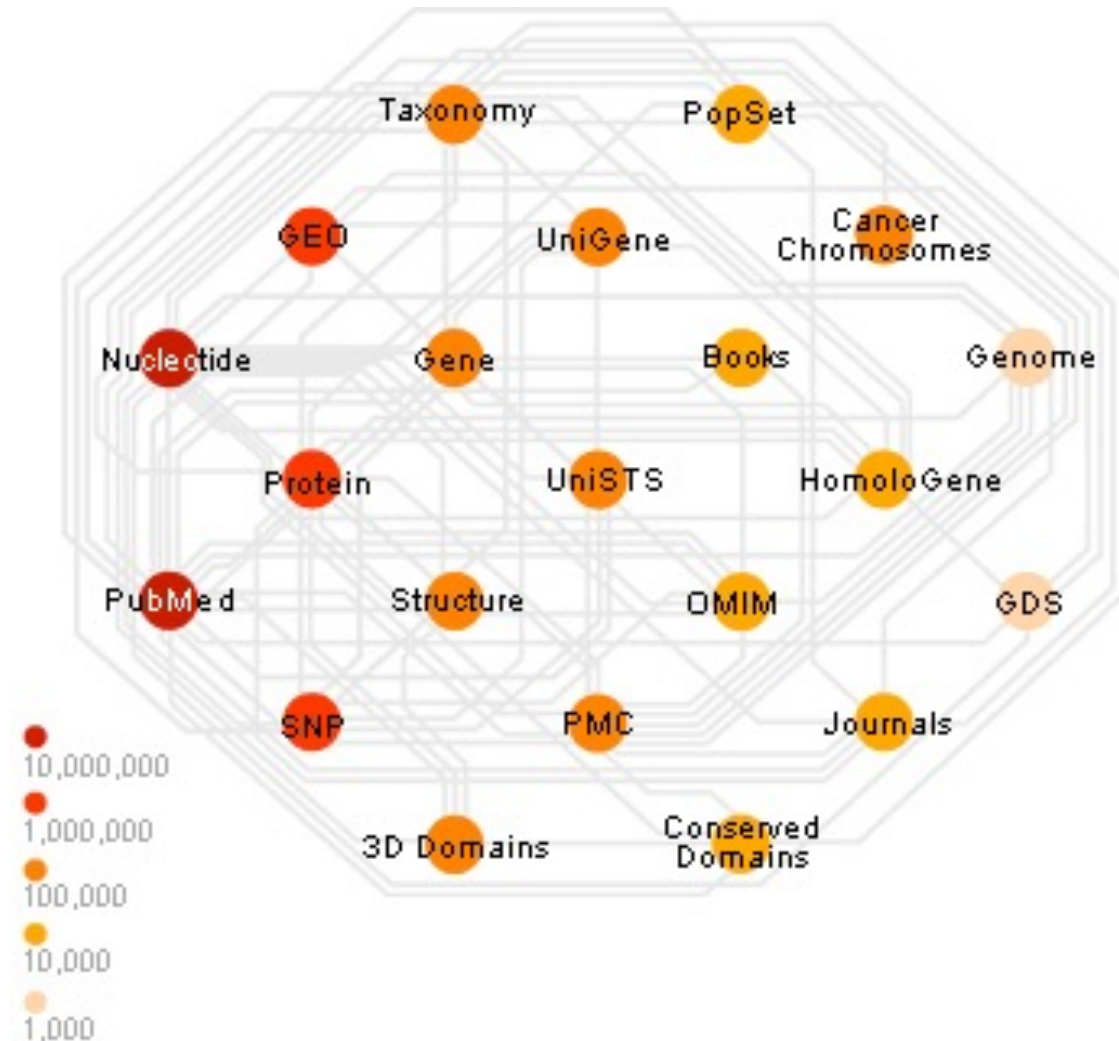


The NCBI system (formerly Entrez) integrates...

- the scientific literature
- DNA and protein sequence databases
- 3D protein structure data
- population study data sets
- assemblies of complete genomes
- many other databases and tools

NCBI database integration

The NCBI implements a search and retrieval system that integrates the NCBI databases



OMIM

OMIM

Limits Advanced

Search

Help



OMIM

OMIM is a comprehensive, authoritative compendium of human genes and genetic phenotypes that is freely available and updated daily. OMIM is authored and edited at the McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University School of Medicine, under the direction of Dr. Ada Hamosh. Its official home is omim.org.

Using OMIM

[Getting Started](#)

[FAQ](#)

OMIM tools

[OMIM API](#)

Related Resources

[ClinVar](#)

[Gene](#)

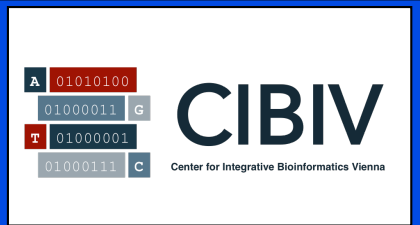
[GTR](#)

[MedGen](#)

OMIM is...

- Online Mendelian Inheritance in Man
- catalog of human genes and genetic disorders
- manually curated

NCBI Bookshelf



NCBI Resources How To

[haschmidt](#) [My NCBI](#) [Sign Out](#)

Bookshelf

Books



Search

[Browse Titles](#) [Advanced](#)

[Help](#)



Bookshelf

Bookshelf provides free online access to books and documents in life science and healthcare. Search, read, and discover.

Using Bookshelf

[Quick Start Guide](#)

[FAQ](#)

Read

[Browse Titles](#)

[New Releases](#)

Participate

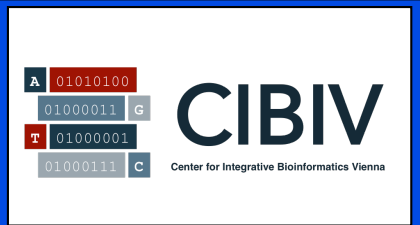
[Authors and Publishers](#)

[How to Apply](#)

Bookshelf is...

- searchable resource of on-line books

NCBI Taxonomy



NCBI Resources How To

haschmidt My NCBI Sign Out

Taxonomy

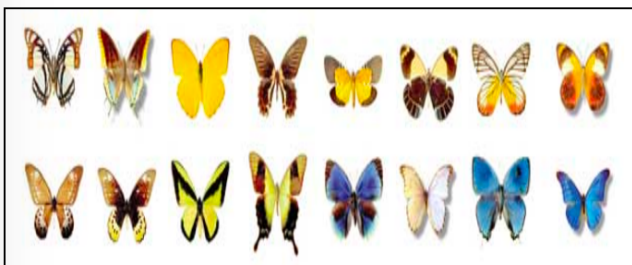
Taxonomy



Search

Limits Advanced

Help



Taxonomy

The Taxonomy Database is a curated classification and nomenclature for all of the organisms in the public sequence databases. This currently represents about 10% of the described species of life on the planet.

Using Taxonomy

[Quick Start Guide](#)

[FAQ](#)

Taxonomy Tools

[Browser](#)

[Common Tree](#)

Other Resources

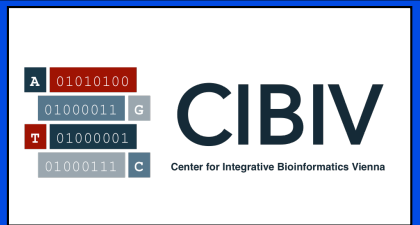
[GenBank](#)

[LinkOut](#)

Taxonomy Browser is...

- browser for the major divisions of living organisms (archaea, bacteria, eukaryota, viruses)
- taxonomy information such as genetic codes
- data sources for each taxon
- molecular data on extinct organisms

Accessing Information



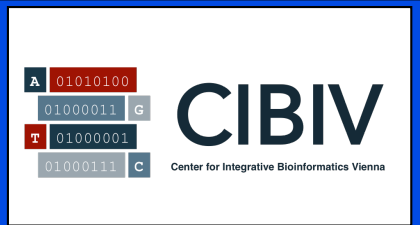
Accession numbers are labels for sequences

NCBI includes databases (such as GenBank) that contain information on DNA, RNA, or protein sequences.

You may want to acquire information beginning with a query such as the **name** of a protein of interest, or the raw nucleotides comprising a DNA **sequence** of interest.

DNA sequences and other molecular data are tagged with **accession numbers** that are used to **uniquely identify** a sequence or other record relevant to molecular data.

Accessing Information



What is an accession number?

An accession number is a **unique label** that is used to **identify** a sequence. It is a string of letters and/or numbers that corresponds to a molecular sequence.

Examples (all for retinol-binding protein, RBP4) from different resources:

- X02775 GenBank genomic DNA sequence
- NT_030059 Genomic contig
- Rs7079946 dbSNP (single nucleotide polymorphism)

- N91759.1 An expressed sequence tag (1 of 170)
- NM_006744 RefSeq DNA sequence (from a transcript)

- NP_007635 RefSeq protein
- AAC02945 GenBank protein
- Q28369 SwissProt protein
- 1KT7 Protein Data Bank structure record

Some ways to access DNA and protein sequences

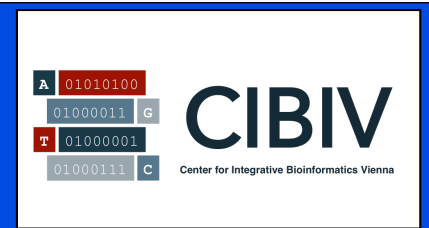
- [1] [NCBI Gene with RefSeq](#)
- [2] European Bioinformatics Institute (EBI) and Ensembl (separate from NCBI)
- [3] ExPASy Sequence Retrieval System (separate from NCBI, Switzerland)

NCBI Gene with RefSeq

NCBI Gene is a great starting point: it collects **key information** on each gene/protein from **major databases**. It covers **all major organisms**.

RefSeq provides a curated, optimal accession number for each DNA (NM_006744) or protein (NP_007635) entry

NCBI Databases



At the NCBI home page enter “rbp4” and hit “Search”



NCBI Resources [v] How To [v] haschmidt My NCBI

NCBI National Center for Biotechnology Information

All Databases [v]

- NCBI Home
- Resource List (A-Z)
- All Resources
- Chemicals & Bioassays
- Data & Software
- DNA & RNA
- Domains & Structures
- Genes & Expression
- Genetics & Medicine
- Genomes & Maps
- Homology
- Literature
- Proteins
- Sequence Analysis
- Taxonomy
- Training & Tutorials

Welcome to NCBI

The National Center for Biotechnology Information advances science and health by providing access to biomedical and genomic information.

[About the NCBI](#) | [Mission](#) | [Organization](#) | [NCBI News & Blog](#)

Submit

Deposit data or manuscripts into NCBI databases



Download

Transfer NCBI data to your computer



Learn

Find help documents, attend a class or watch a tutorial



Develop

Use NCBI APIs and

Analyze

Identify an NCBI tool

Research

Explore NCBI research

Popular Resources

[PubMed](#)

[Bookshelf](#)

[PubMed Central](#)

[BLAST](#)

[Nucleotide](#)

[Genome](#)

[SNP](#)

[Gene](#)

[Protein](#)

[PubChem](#)

NCBI News & Blog

Posters, Presentations, and a Book of Abstracts from NCBI at ASM Microbe 2019

From June 20-24, 2019, NCBI

Entrez Databases: (2006)



Entrez, The Life Sciences Search Engine.

HOME SEARCH SITE MAP PubMed Entrez Human Genome GenBank Map Viewer BLAST

Search across databases

GO

CLEAR

Help

| | | | | | | | |
|------|--|---|---|------|--|---|---|
| 11 | | PubMed: biomedical literature citations and abstracts | ? | none | | Books: online books | ? |
| 2 | | PubMed Central: free, full text journal articles | ? | 5 | | OMIM: Online Mendelian Inheritance in Man | ? |
| | | | | none | | Site Search: NCBI web and FTP sites | ? |
| 40 | | Nucleotide: sequence database (GenBank) | ? | 7 | | UniGene: gene-oriented clusters of transcript sequences | ? |
| 25 | | Protein: sequence database | ? | none | | CDD: conserved protein domain database | ? |
| 1 | | Genome: whole genome sequences | ? | none | | 3D Domains: domains from Entrez Structure | ? |
| none | | Structure: three-dimensional macromolecular structures | ? | 9 | | UniSTS: markers and mapping data | ? |
| none | | Taxonomy: organisms in GenBank | ? | 1 | | PopSet: population study data sets | ? |
| 86 | | SNP: single nucleotide polymorphism | ? | 489 | | GEO Profiles: expression and molecular abundance profiles | ? |
| 8 | | Gene: gene-centered information | ? | none | | GEO DataSets: experimental sets of GEO data | ? |
| 20 | | HomoloGene: Eukaryotic homology groups | ? | none | | Cancer Chromosomes: cytogenetic databases | ? |
| none | | Journals: detailed information about the journals indexed in PubMed and other Entrez databases | ? | none | | MeSH: detailed information about NLM's controlled vocabulary | ? |

Entrez Databases: (2010)



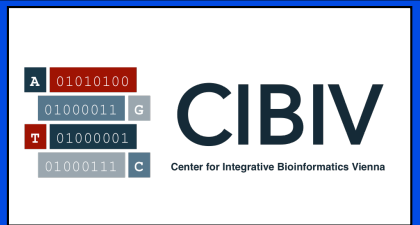
Search across databases

[Help](#)

Result counts displayed in gray indicate one or more terms not found

| | |
|--|--|
| 317 PubMed: biomedical literature citations and abstracts | Books: online books |
| 213 PubMed Central: free, full text journal articles | 114 Images: images from full text resources at NCBI |
| none Site Search: NCBI web and FTP sites | 7 OMIM: online Mendelian Inheritance in Man |
| 171 Nucleotide: Core subset of nucleotide sequence records | none dbGaP: genotype and phenotype |
| 52 EST: Expressed Sequence Tag records | 14 UniGene: gene-oriented clusters of transcript sequences |
| 3 GSS: Genome Survey Sequence records | none CDD: conserved protein domain database |
| 141 Protein: sequence database | 15 UniSTS: markers and mapping data |
| 22 Genome: whole genome sequences | 1 PopSet: population study data sets |
| 4 Structure: three-dimensional macromolecular structures | 5386 GEO Profiles: expression and molecular abundance profiles |
| none Taxonomy: organisms in GenBank | 1 GEO DataSets: experimental sets of GEO data |
| 590 SNP: single nucleotide polymorphism | none Epigenomics: Epigenetic maps and data sets |
| none dbVar: Genomic structural variation | none Cancer Chromosomes: cytogenetic databases |
| 88 Gene: gene-centered information | 2 PubChem BioAssay: bioactivity screens of chemical substances |
| none SRA: Sequence Read Archive | none PubChem Compound: unique small molecule chemical structures |
| none BioSystems: Pathways and systems of interacting molecules | 10 PubChem Substance: deposited chemical substance records |
| 3 HomoloGene: eukaryotic homology groups | none Protein Clusters: a collection of related protein sequences |
| 94 GENSAT: gene expression atlas of mouse central nervous system | 18 Peptidome: MS/MS proteomic experiments |

NCBI Databases: (2013)



Literature

| | | | |
|---------------------|--|--------------------|--|
| 534 | PubMed : scientific & medical abstracts/citations | 5 | MeSH : ontology used for PubMed indexing |
| 632 | PubMed Central : full-text journal articles | 3 | Books : books and reports |
| 0 | NLM Catalog : books, journals and more in the NLM Collections | 11 | Site Search : NCBI web and FTP site index |

Health

| | | | |
|--------------------|---|-------------------|--|
| 3 | PubMed Health : clinical effectiveness, disease and drug reports | 6 | ClinVar : human variations of clinical significance |
| 1 | MedGen : medical genetics literature and links | 9 | OMIM : online mendelian inheritance in man |
| 3 | GTR : genetic testing registry | 0 | OMIA : online mendelian inheritance in animals |
| 17 | dbGaP : genotype/phenotype interaction studies | | |

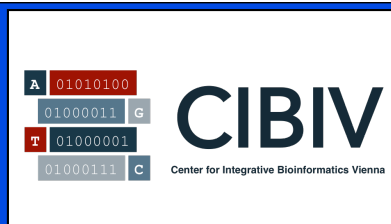
Organisms

| | |
|-------------------|---|
| 0 | Taxonomy : taxonomic classification and nomenclature catalog |
|-------------------|---|

Nucleotide Sequences

| | | | |
|---------------------|---|---------------------|--|
| 392 | Nucleotide : DNA and RNA sequences | 0 | SRA : high-throughput DNA and RNA sequence read archive |
| 3 | GSS : genome survey sequences | 2 | PopSet : sequence sets from phylogenetic and population studies |
| 52 | EST : expressed sequence tag sequences | 147 | Probe : sequence-based probes and primers |

NCBI Databases: (2016)



Literature

| | | |
|-----------------------|-------|---|
| Books | 8 | books and reports |
| MeSH | 5 | ontology used for PubMed indexing |
| NLM Catalog | 0 | books, journals and more in the NLM Collections |
| PubMed | 769 | scientific & medical abstracts/citations |
| PubMed Central | 1,265 | full-text journal articles |

Health

| | | |
|----------------------|----|--|
| ClinVar | 10 | human variations of clinical significance |
| dbGaP | 20 | genotype/phenotype interaction studies |
| GTR | 16 | genetic testing registry |
| MedGen | 3 | medical genetics literature and links |
| OMIM | 11 | online mendelian inheritance in man |
| PubMed Health | 3 | clinical effectiveness, disease and drug reports |

Genomes

| | | |
|-------------------|-------|---|
| Assembly | 0 | genome assembly information |
| BioProject | 2 | biological projects providing data to NCBI |
| BioSample | 221 | descriptions of biological source materials |
| Clone | 1,079 | genomic and cDNA clones |
| dbVar | 73 | genome structural variation studies |
| Genome | 13 | genome sequencing projects by organism |
| GSS | 3 | genome survey sequences |
| Nucleotide | 794 | DNA and RNA sequences |

Genes

| | | |
|---------------------|-------|--|
| EST | 54 | expressed sequence tag sequences |
| Gene | 273 | collected information about gene loci |
| GEO DataSets | 234 | functional genomics studies |
| GEO Profiles | 8,061 | gene expression and molecular abundance profiles |
| HomoloGene | 3 | homologous gene sets for selected organisms |
| PopSet | 2 | sequence sets from phylogenetic and population studies |
| UniGene | 17 | clusters of expressed transcripts |

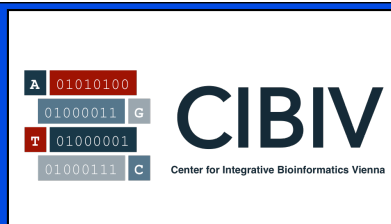
Proteins

| | | |
|--------------------------|-----|---|
| Conserved Domains | 0 | conserved protein domains |
| Protein | 395 | protein sequences |
| Protein Clusters | 0 | sequence similarity-based protein clusters |
| Structure | 36 | experimentally-determined biomolecular structures |

Chemicals

| | | |
|-------------------------|-----|--|
| BioSystems | 406 | molecular pathways with links to genes, proteins and chemicals |
| PubChem BioAssay | 56 | bioactivity screening studies |
| PubChem Compound | 0 | chemical information with structures, information and links |
| PubChem | 102 | deposited substance and chemical |

NCBI Databases: (2019)



Literature

| | |
|----------------|-------|
| Bookshelf | 9 |
| MeSH | 5 |
| NLM Catalog | 2 |
| PubMed | 977 |
| PubMed Central | 2,034 |

Proteins

| | |
|--------------------------|-----|
| Conserved Domains | 0 |
| Identical Protein Groups | 14 |
| Protein | 565 |
| Protein Clusters | 0 |
| Sparcle | 1 |
| Structure | 44 |

Genes

| | |
|--------------|-------|
| Gene | 380 |
| GEO DataSets | 2,073 |
| GEO Profiles | 9,450 |
| HomoloGene | 3 |
| PopSet | 2 |
| UniGene | 17 |

Genomes

| | |
|----------------|-------|
| Assembly | 0 |
| BioCollections | 0 |
| BioProject | 2 |
| BioSample | 2,065 |
| Genome | 60 |
| Nucleotide | 1,068 |
| Probe | 153 |
| SRA | 3,756 |
| Taxonomy | 0 |

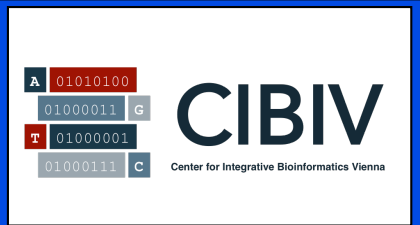
Genetics

| | |
|---------|-------|
| ClinVar | 29 |
| dbGaP | 0 |
| dbSNP | 6,488 |
| dbVar | 234 |
| GTR | 33 |
| MedGen | 4 |
| OMIM | 11 |

Chemicals

| | |
|-------------------|-----|
| BioSystems | 418 |
| PubChem BioAssay | 136 |
| PubChem Compound | 0 |
| PubChem Substance | 122 |

NCBI Databases: (2022)



Search NCBI

rbp4



Search

Results found in 29 databases

GENE

Was this helpful?



RBP4 – retinol binding protein 4

Homo sapiens (human)

Also known as: MCOPCB10, RDCCAS

Gene ID: 5950

[RefSeq transcripts](#) (3) [RefSeq proteins](#) (3) [RefSeqGene](#) (1) [PubMed](#) (382)

[Orthologs](#)

[Genome Data Viewer](#)

[BLAST](#)

[Download](#)

RefSeq Sequences



RefSeq transcripts

RBP4 – 3 of 3 transcripts

| Transcript | Isoform | Len (nt) |
|----------------|-------------|----------|
| NM_001323518.2 | b | 1,009 |
| NM_006744.4 | a precursor | 1,070 |
| NM_001323517.1 | a precursor | 1,015 |

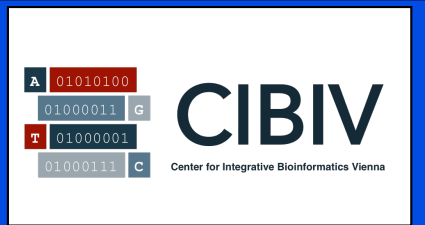
[View full table](#)

Literature

Genes

Proteins

NCBI Databases: (2022)



| Literature | |
|----------------|-------|
| Bookshelf | 17 |
| MeSH | 5 |
| NLM Catalog | 2 |
| PubMed | 1,365 |
| PubMed Central | 3,897 |

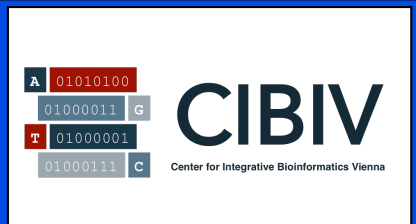
| Genes | |
|--------------|-------|
| Gene | 579 |
| GEO DataSets | 6,387 |
| GEO Profiles | 9,450 |
| HomoloGene | 3 |
| PopSet | 2 |

| Proteins | |
|--------------------------|-----|
| Conserved Domains | 2 |
| Identical Protein Groups | 19 |
| Protein | 889 |
| Protein Family Models | 0 |
| Structure | 35 |

| Genomes | |
|----------------|-------|
| Assembly | 0 |
| BioCollections | 0 |
| BioProject | 7 |
| BioSample | 6,367 |
| Genome | 239 |
| Nucleotide | 1,869 |
| SRA | 8,070 |
| Taxonomy | 0 |

| Clinical | |
|--------------------|-------|
| ClinicalTrials.gov | 59 |
| ClinVar | 137 |
| dbGaP | 4 |
| dbSNP | 4,533 |
| dbVar | 249 |
| GTR | 47 |
| MedGen | 2 |
| OMIM | 11 |

| PubChem | |
|------------|-----|
| BioAssays | 203 |
| Compounds | 2 |
| Pathways | 0 |
| Substances | 79 |



Gene [Create RSS](#) [Save search](#) [Advanced](#) [Help](#)

- Gene sources**
Genomic
- Categories**
Alternatively spliced
Annotated genes
Protein-coding
- Sequence content**
CCDS
Ensembl
RefSeq
RefSeqGene
- Status**
 Current
- [Clear all](#)
[Show additional filters](#)

Tabular 20 per page Sort by Relevance Send to:

[Hide sidebar >>](#)

Filters: [Manage Filters](#)

GENE Was this helpful?

[RBP4 – retinol binding protein 4](#)

[Homo sapiens \(human\)](#)

Also known as: MCOPCB10, RDCCAS

Gene ID: 5950

[RefSeq transcripts](#) (3) [RefSeq proteins](#) (3) [RefSeqGene](#) (1) [PubMed](#) (382)

RefSeq Sequences +

Results by taxon

- Top Organisms** [\[Tree\]](#)
- Homo sapiens (66)
 - Drosophila melanogaster (27)
 - Mus musculus (18)
 - Sus scrofa (5)
 - Rattus norvegicus (2)
 - All other taxa (461)
- [More...](#)

Find related data

Database:

Search details

rbp4[All Fields] AND alive[prop]

[See more...](#)

Search results

Items: 1 to 20 of 579

<< First < Prev Page of 29 [Next >](#) [Last >>](#)

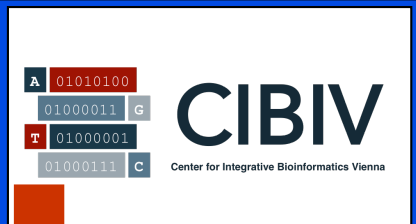
[See also 10 discontinued or replaced items.](#)

| Name/Gene ID | Description | Location | Aliases | MIM |
|--|--|---|------------------|--------|
| <input type="checkbox"/> RBP4 ID: 5950 | retinol binding protein 4 [<i>Homo sapiens</i> (human)] | Chromosome 10, NC_000010.11 (93591694..93601744, complement) | MCOPCB10, RDCCAS | 180250 |
| <input type="checkbox"/> Rbp4 ID: 19662 | retinol binding protein 4, plasma [<i>Mus musculus</i> (house mouse)] | Chromosome 19, NC_000085.7 (38105068..38113769, complement) | Rbp-4 | |

Recent activity

Your browsing activity is temporarily unavailable.

NCBI Gene (refine search)



Gene [Create RSS](#) [Save search](#) [Advanced](#) [Help](#)

- Gene sources
 - Genomic
 - Categories
 - Alternatively spliced
 - Annotated genes
 - Protein-coding
 - Sequence content
 - CCDS
 - Ensembl
 - RefSeq
 - RefSeqGene
 - Status
 - Current
- [Clear all](#)
[Show additional filters](#)

Tabular 20 per page Sort by Relevance Send to: [Hide sidebar >>](#)

GENE Was this helpful?

[RBP4 – retinol binding protein 4](#)

[Homo sapiens \(human\)](#)

Also known as: MCOPCB10, RDCCAS

Gene ID: 5950

[RefSeq transcripts \(3\)](#) [RefSeq proteins \(3\)](#) [RefSeqGene \(1\)](#) [PubMed \(382\)](#)

RefSeq Sequences

Filters: [Manage Filters](#)

- Results by taxon**
- [Top Organisms \[Tree\]](#)
- [Homo sapiens \(66\)](#)
 - [Drosophila melanogaster \(27\)](#)
 - [Mus musculus \(18\)](#)
 - [Sus scrofa \(5\)](#)
 - [Rattus norvegicus \(2\)](#)
 - [All other taxa \(461\)](#)
- [More...](#)

Find related data

Database:

Search details

rbp4[All Fields] AND alive[prop]

[See more...](#)

Search results

Items: 1 to 20 of 579 << First < Prev Page 1 of 29 Next > Last >>

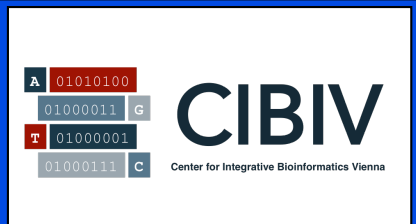
[See also 10 discontinued or replaced items.](#)

| Name/Gene ID | Description | Location | Aliases | MIM |
|--|--|--|------------------|--------|
| <input type="checkbox"/> RBP4 ID: 5950 | retinol binding protein 4 [<i>Homo sapiens</i> (human)] | Chromosome 10, NC_000010.11 (93591694..93601744, complement) | MCOPCB10, RDCCAS | 180250 |
| <input type="checkbox"/> Rbp4 ID: 19662 | retinol binding protein 4, plasma [<i>Mus musculus</i> (house mouse)] | Chromosome 19, NC_000085.7 (38105068..38113769, complement) | Rbp-4 | |

Recent activity

Your browsing activity is temporarily unavailable.

Filters reduce the number of entries



Gene [Create RSS](#) [Save search](#) [Advanced](#) [Help](#)

Gene sources: Genomic Send to:

- Categories
- Alternatively spliced
 - Annotated genes
 - Protein-coding

- Sequence content
- CCDS
 - Ensembl
 - RefSeq
 - RefSeqGene

Status

Current

[Clear all](#)

[Show additional filters](#)

Tabular Sort by Relevance

Search results

Items: 1 to 20 of 66

Showing Current items.

<< First < Prev Page 1 of 4 Next > Last >>

| Name/Gene ID | Description | Location | Aliases | MIM |
|---|--|--|--|--------|
| <input type="checkbox"/> RBP4 ID: 5950 | retinol binding protein 4 [<i>Homo sapiens</i> (human)] | Chromosome 10, NC_000010.11 (93591694..93601744, complement) | MCOPCB10, RDCCAS | 180250 |
| <input type="checkbox"/> POLR2D ID: 5433 | RNA polymerase II subunit D [<i>Homo sapiens</i> (human)] | Chromosome 2, NC_000002.12 (127843553..127858155, complement) | HSRBP4, HSRPB4, RBP4, RPB16, RPB4 | 606017 |
| <input type="checkbox"/> ADIPOQ ID: 9370 | adiponectin, C1Q and collagen domain containing [<i>Homo sapiens</i> (human)] | Chromosome 3, NC_000003.12 (186842710..186858463) | ACDC, ACRP30, ADIPQTL1, ADPN, APM-1, APM1, GBP28 | 605441 |
| <input type="checkbox"/> LEP ID: 3952 | leptin [<i>Homo sapiens</i> (human)] | Chromosome 7, NC_000007.14 (128241278..128257629) | LEPD, OB, OBS | 164160 |
| <input type="checkbox"/> ACE ID: 1636 | angiotensin I converting enzyme [<i>Homo sapiens</i> (human)] | Chromosome 17, NC_000017.11 (63477061..63498373) | ACE1, CD143, DCP, DCP1 | 106180 |
| <input type="checkbox"/> CRP ID: 1401 | C-reactive protein [<i>Homo</i> <i>sapiens</i> (human)] | Chromosome 1, NC_000001.11 (159712289..159714589, complement) | PTX1 | 123260 |
| <input type="checkbox"/> IL1B ID: 3553 | interleukin 1 beta [<i>Homo</i> <i>sapiens</i> (human)] | Chromosome 2, NC_000002.12 (112829751..112836779, complement) | IL-1, IL1-BETA, IL1F2, IL1beta | 147720 |

Filters: [Manage Filters](#)

Find related data

Database:

Search details

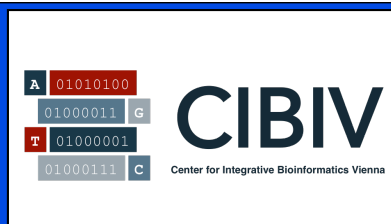
rbp4[All Fields] AND "Homo sapiens"
[porgn] AND alive[prop]

[See more...](#)

Recent activity

Your browsing activity is temporarily unavailable.

Gene entry (summary)



RBP4 retinol binding protein 4 [*Homo sapiens* (human)]

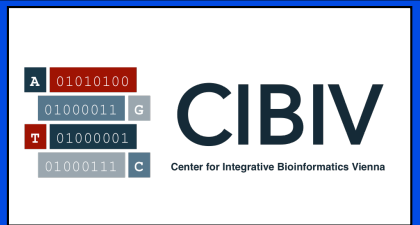
Download Datasets

Gene ID: 5950, updated on 4-Dec-2022

Summary

| | |
|---------------------------|---|
| Official Symbol | RBP4 provided by HGNC |
| Official Full Name | retinol binding protein 4 provided by HGNC |
| Primary source | HGNC:HGNC:9922 |
| See related | Ensembl:ENSG00000138207 MIM:180250 ; AllianceGenome:HGNC:9922 |
| Gene type | protein coding |
| RefSeq status | REVIEWED |
| Organism | Homo sapiens |
| Lineage | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini; Catarrhini; Hominidae; Homo |
| Also known as | RDCCAS; MCOPCB10 |
| Summary | This protein belongs to the lipocalin family and is the specific carrier for retinol (vitamin A alcohol) in the blood. It delivers retinol from the liver stores to the peripheral tissues. In plasma, the RBP-retinol complex interacts with transthyretin which prevents its loss by filtration through the kidney glomeruli. A deficiency of vitamin A blocks secretion of the binding protein posttranslationally and results in defective delivery and supply to the epidermal cells. [provided by RefSeq, Jul 2008] |
| Expression | Biased expression in liver (RPKM 2545.8) and fat (RPKM 387.7) See more |
| Orthologs | mouse all |
| NEW | Try the new Gene table |
| | Try the new Transcript table |

Gene entry (genomic context)



Genomic context

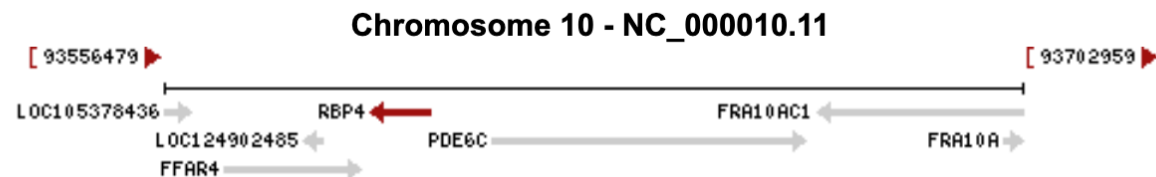


See RBP4 in [Genome Data Viewer](#)

Location: 10q23.33

Exon count: 8

| Annotation release | Status | Assembly | Chr | Location |
|---------------------|-------------------|--|-----|---|
| 110 | current | GRCh38.p14 (GCF_000001405.40) | 10 | NC_000010.11 (93591694..93601744, complement) |
| 110 | current | T2T-CHM13v2.0 (GCF_009914755.1) | 10 | NC_060934.1 (94471646..94481695, complement) |
| 105.20220307 | previous assembly | GRCh37.p13 (GCF_000001405.25) | 10 | NC_000010.10 (95351451..95361501, complement) |



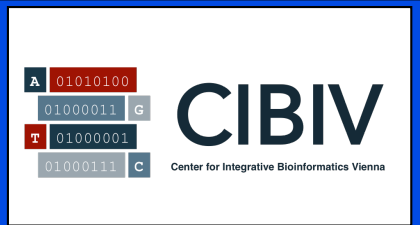
Genomic regions, transcripts, and products



Go to [reference sequence details](#)

Genomic Sequence: ▼

Gene entry (genomic region)

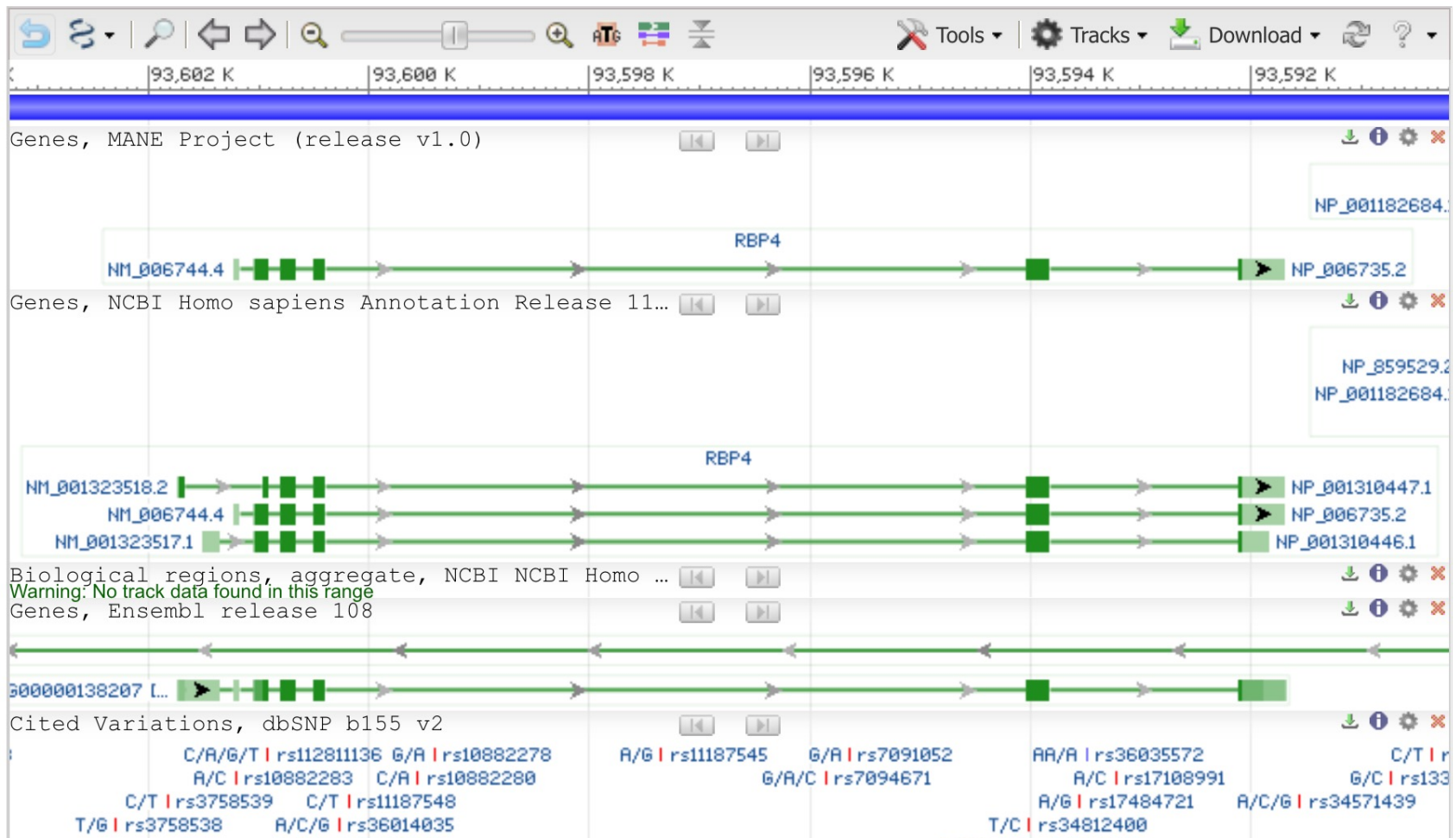


Genomic regions, transcripts, and products

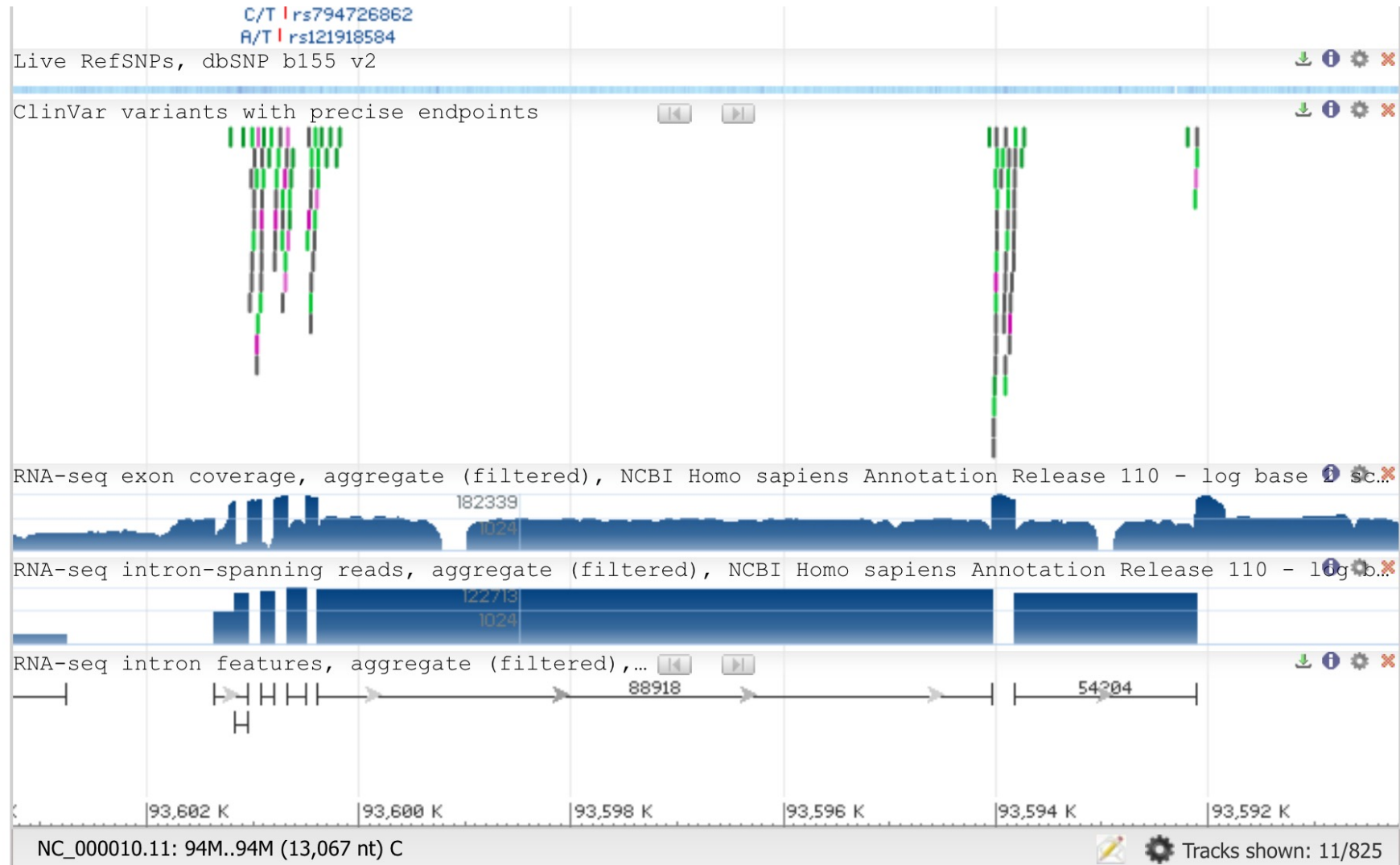
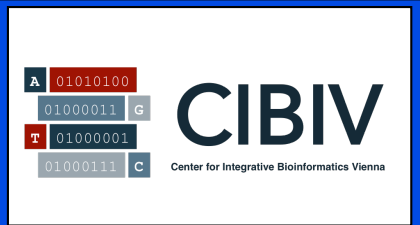
[Go to reference sequence details](#)

Genomic Sequence: NC_000010.11 Chromosome 10 Reference GRCh38.p14 Primary Assembly

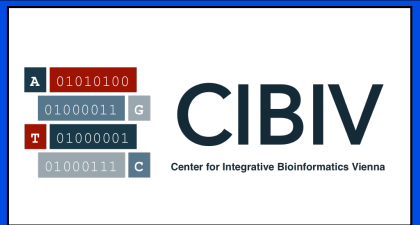
[Go to nucleotide: Graphics](#) [FASTA](#) [GenBank](#)



Gene entry (genomic region, cont'd)



Gene entry (expression)

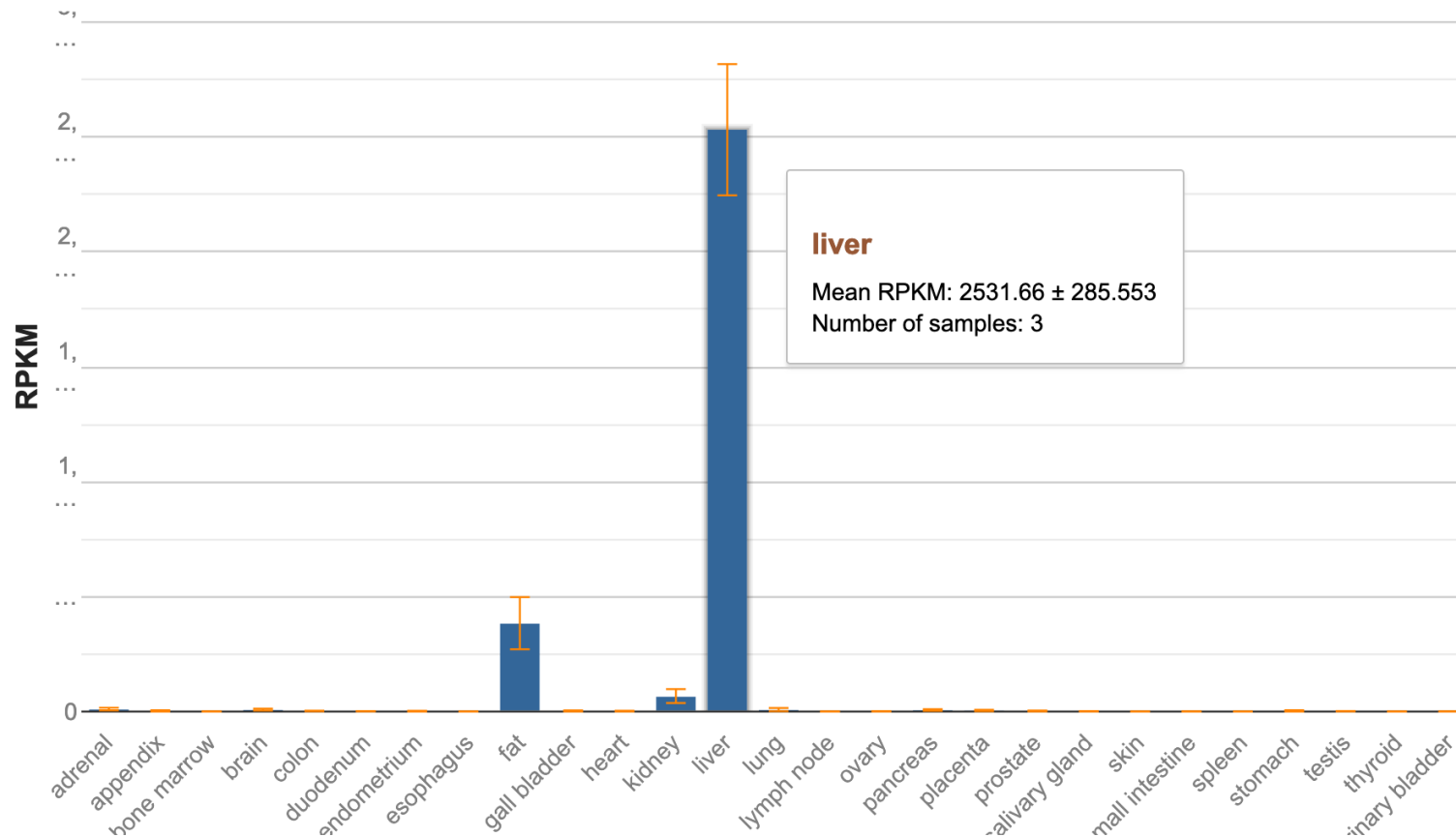


Expression

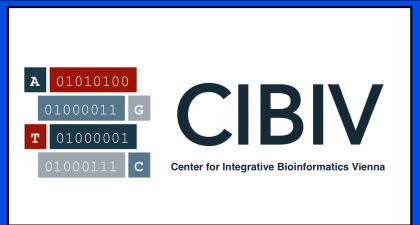
[See details](#)

HPA RNA-seq normal tissues

- Project title: HPA RNA-seq normal tissues
- Description: RNA-seq was performed of tissue samples from 95 human individuals representing 27 different tissues in order to determine tissue-specificity of all protein-coding genes
- BioProject: [PRJEB4337](#)
- Publication: [PMID 24309898](#)
- Analysis date: Wed Apr 4 07:08:55 2018



Gene entry (overview)



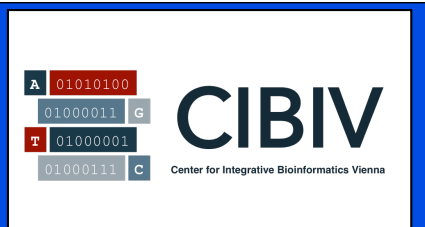
RBP4 retinol binding protein 4 [*Homo sapiens* (human)]

[Download Datasets](#)

Gene ID: 5950, updated on 4-Dec-2022

- Summary ⬆️ ?
- Genomic context ⬆️ ?
- Genomic regions, transcripts, and products ⬆️ ?
- Expression ⬆️ ?
- Bibliography ⬆️ ?
- Phenotypes ⬆️ ?
- Variation ⬆️ ?
- Pathways from PubChem ⬆️
- Interactions ⬆️ ?
- General gene information ⬆️ ?
- General protein information ⬆️ ?
- NCBI Reference Sequences (RefSeq) ⬆️ ?
- Related sequences ⬆️ ?
- Additional links ⬆️ ?

Gene entry (protein via RefSeq part)



mRNA and Protein(s)

1. [NM_001323517.1](#) → [NP_001310446.1](#) retinol-binding protein 4 isoform a precursor

Status: REVIEWED

| | | |
|---|---|--|
| Description | Transcript Variant: This variant (2) and variant 1 both encode isoform a. | |
| Source sequence(s) | AL356214 | |
| Consensus CDS | CCDS31249.1 | |
| UniProtKB/Swiss-Prot | P02753 | |
| Related | ENSP00000360522.1 , ENST00000371467.5 | |
| Conserved Domains (1) summary | | |
| | pfam00061 Location:39 → 177 | Lipocalin; Lipocalin / cytosolic fatty-acid binding protein family |

2. [NM_001323518.1](#) → [NP_001310447.1](#) retinol-binding protein 4 isoform b

Status: REVIEWED

| | | |
|---|---|--|
| Source sequence(s) | AL356214 , BC020633 , BG565176 , BI712834 | |
| Consensus CDS | CCDS81488.1 | |
| UniProtKB/Swiss-Prot | P02753 | |
| UniProtKB/TrEMBL | Q5VY30 | |
| Related | ENSP00000360524.2 , ENST00000371469.2 | |
| Conserved Domains (1) summary | | |
| | pfam00061 Location:37 → 175 | Lipocalin; Lipocalin / cytosolic fatty-acid binding protein family |

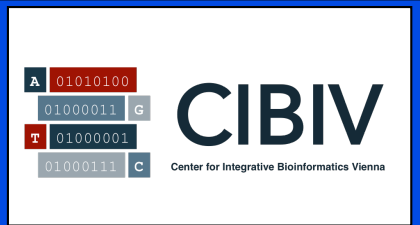
3. [NM_006744.4](#) → [NP_006735.2](#) retinol-binding protein 4 isoform a precursor

[See identical proteins and their annotated locations for NP_006735.2](#)

Status: REVIEWED

| | | |
|--------------------|---|--|
| Description | Transcript Variant: This variant (1) and variant 2 both encode isoform a. | |
| Source sequence(s) | AL356214 , BC020633 , BX495987 , X00129 | |
| Consensus CDS | CCDS31249.1 | |

Protein entry (top)



Protein

Protein



Search

Advanced

Help

GenPept ▾

Send to: ▾

Change region shown ▾

Customize view ▾

retinol-binding protein 4 isoform a precursor [Homo sapiens]

NCBI Reference Sequence: NP_001310446.1

[Identical Proteins](#) [FASTA](#) [Graphics](#)

Go to: ▾

LOCUS NP_001310446 201 aa linear PRI 28-MAY-2019
DEFINITION retinol-binding protein 4 isoform a precursor [Homo sapiens].
ACCESSION NP_001310446
VERSION NP_001310446.1
DBSOURCE REFSEQ: accession [NM_001323517.1](#)
KEYWORDS RefSeq.
SOURCE Homo sapiens (human)
ORGANISM [Homo sapiens](#)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini; Catarrhini; Hominidae; Homo.
REFERENCE 1 (residues 1 to 201)
AUTHORS Aref-Eshghi E, Hurley O, Sun G, Simms A, Godwin M, Duke P, Araee M, Mahdavian M and Asghari S.
TITLE Genetic associations in community context: a mixed model approach identifies a functional variant in the RBP4 gene associated with HDL-C dyslipidemia
JOURNAL BMC Med. Genet. 19 (1), 205 (2018)
PUBMED [30497399](#)
REMARK GeneRIF: A mixed model approach identifies a functional variant in the RBP4 gene associated with HDL-C dyslipidemia the in Newfoundland and Labrador population.
Publication Status: Online-Only
REFERENCE 2 (residues 1 to 201)
AUTHORS Wang H, Zhou P, Zou D, Liu Y, Lu X and Liu Z.
TITLE The role of retinol-binding protein 4 and its relationship with sex hormones in coronary artery disease
JOURNAL Biochem. Biophys. Res. Commun. 506 (1), 204-210 (2018)
PUBMED [30342852](#)
REMARK GeneRIF: RBP4 levels were significantly decreased and positively

Analyze this sequence ▾

Run BLAST

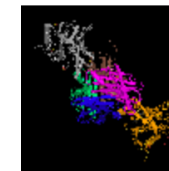
Identify Conserved Domains

Highlight Sequence Features

Find in this Sequence

Show in Genome Data Viewer

Protein 3D Structure ▾



Complex of TTR and RBP4 and Oleic Acid
PDB: 2WQA
Source: Homo sapiens
Method: X-ray Diffraction

Resolution: 2.85 Å

[See all 22 structures...](#)

Articles about the RBP4 gene ▾

Genetic associations in community context: a mixed r [BMC Med Genet. 2018]

The role of retinol-binding protein 4 and its rel: [Biochem Biophys Res Commun. 2018]

Elevated RBP4 plasma levels were

Protein entry (bottom: features, sequence)

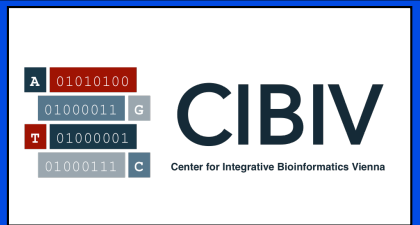


```
FEATURES
    source
        1..201
        /organism="Homo sapiens"
        /db_xref="taxon:9606"
        /chromosome="10"
        /map="10q23.33"
    Protein
        1..201
        /product="retinol-binding protein 4 isoform a precursor"
        /note="retinol-binding protein 4, interstitial; RBP; PRBP;
        plasma retinol-binding protein; retinol binding protein 4,
        plasma"
        /calculated_mol_wt=21072
    sig_peptide
        1..18
        /inference="COORDINATES: ab initio prediction:SignalP:4.0"
        /calculated_mol_wt=1956
    mat_peptide
        19..201
        /product="retinol-binding protein 4 isoform a"
        /calculated_mol_wt=21072
    Region
        39..177
        /region_name="Lipocalin"
        /note="Lipocalin / cytosolic fatty-acid binding protein
        family; pfam00061"
        /db_xref="CDD:306552"
    Site
        139
        /site_type="methylation"
        /experiment="experimental evidence, no additional details
        recorded"
        /note="Omega-N-methylarginine.
        {ECO:0000250|UniProtKB:Q00724}; propagated from
        UniProtKB/Swiss-Prot (P02753.3)"
    CDS
        1..201
        /gene="RBP4"
        /gene_synonym="MCOPCB10; RDCCAS"
        /coded_by="NM_001323517.1:171..776"
        /note="isoform a precursor is encoded by transcript
        variant 2"
        /db_xref="CCDS:CCDS31249.1"
        /db_xref="GeneID:5950"
        /db_xref="HGNC:HGNC:9922"
        /db_xref="MIM:180250"
```

ORIGIN

```
1 mkvwwalllll aalgsgraer dcrvssfrvk enfdkarfsg twyamakkdp eglflqdniv
61 aefsvdetgq msatakgrvr llnnwdvcad mvgftftded pakfkmkywg vasflqkgnd
121 dhwivdtdyd tyavqyscrl lnldgtcads ysfvfsrdpn glppeaqkiv rqrqeelcla
181 rqyrlivhng ycdgrsernl l
..
```


Protein entry (top)



Protein

Protein



Search

Advanced

Help

GenPept ▾

Send to: ▾

Change region shown ▾

Customize view ▾

retinol-binding protein 4 isoform a precursor [Homo sapiens]

NCBI Reference Sequence: NP_001310446.1

[Identical Proteins](#) [FASTA](#) [Graphics](#)

Go to: ▾

LOCUS NP_001310446 201 aa linear PRI 28-MAY-2019
DEFINITION retinol-binding protein 4 isoform a precursor [Homo sapiens].
ACCESSION NP_001310446
VERSION NP_001310446.1
DBSOURCE REFSEQ: accession [NM_001323517.1](#)
KEYWORDS RefSeq.
SOURCE Homo sapiens (human)
ORGANISM [Homo sapiens](#)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini; Catarrhini; Hominidae; Homo.
REFERENCE 1 (residues 1 to 201)
AUTHORS Aref-Eshghi E, Hurley O, Sun G, Simms A, Godwin M, Duke P, Araee M, Mahdavian M and Asghari S.
TITLE Genetic associations in community context: a mixed model approach identifies a functional variant in the RBP4 gene associated with HDL-C dyslipidemia
JOURNAL BMC Med. Genet. 19 (1), 205 (2018)
PUBMED [30497399](#)
REMARK GeneRIF: A mixed model approach identifies a functional variant in the RBP4 gene associated with HDL-C dyslipidemia the in Newfoundland and Labrador population.
Publication Status: Online-Only
REFERENCE 2 (residues 1 to 201)
AUTHORS Wang H, Zhou P, Zou D, Liu Y, Lu X and Liu Z.
TITLE The role of retinol-binding protein 4 and its relationship with sex hormones in coronary artery disease
JOURNAL Biochem. Biophys. Res. Commun. 506 (1), 204-210 (2018)
PUBMED [30342852](#)
REMARK GeneRIF: RBP4 levels were significantly decreased and positively

Analyze this sequence ▾

Run BLAST

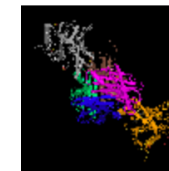
Identify Conserved Domains

Highlight Sequence Features

Find in this Sequence

Show in Genome Data Viewer

Protein 3D Structure ▾



Complex of TTR and RBP4 and Oleic Acid
PDB: 2WQA
Source: Homo sapiens
Method: X-ray Diffraction

Resolution: 2.85 Å

[See all 22 structures...](#)

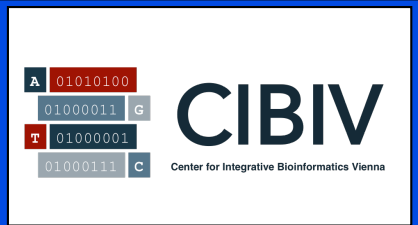
Articles about the RBP4 gene ▾

Genetic associations in community context: a mixed r [BMC Med Genet. 2018]

The role of retinol-binding protein 4 and its rel: [Biochem Biophys Res Commun. 2018]

Elevated RBP4 plasma levels were

Protein entry (sequence)



FASTA ▾

Send to: ▾

retinol-binding protein 4 isoform a precursor [Homo sapiens]

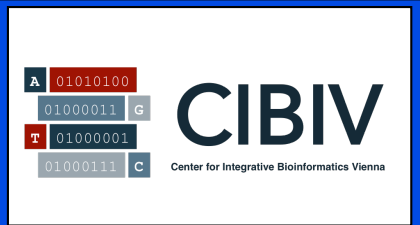
NCBI Reference Sequence: NP_001310446.1

[GenPept](#) [Identical Proteins](#) [Graphics](#)

```
>gi|1021087280|ref|NP_001310446.1| retinol-binding protein 4 isoform a precursor  
[Homo sapiens]
```

```
MKWVWALLLLAALGSGRAERDCRVSSFRVKENFDKARFSGTWYAMAKKDPEGLFLQDNIVAEFVSDETGQ  
MSATAKGRVRLNNDVDCADMVGTFTDTEDPAKFKMKYWGVASFQKGNDDHWIVDTDYDTYAVQYSCRL  
LNLDTGTCADSYFVFSRDPNGLPPEAQKIVRQRQEELCLARQYRLIVHNGYCDGRSERNLL
```

Protein entry (download, FASTA format)



FASTA ▾

retinol-binding protein 4 isoform a precursor [Homo

NCBI Reference Sequence: NP_001310446.1

[GenPept](#) [Identical Proteins](#) [Graphics](#)

```
>gi|1021087280|ref|NP_001310446.1| retinol-binding protein 4 isoform  
[Homo sapiens]
```

```
MKWVWALLLLAALGSGRAERDCRVSSFRVKENFDKARFSGTWYAMAKKDPEGLFLQDNIVAEFSVDETC  
MSATAKGRVRLNNDVVCADMVGTFTDTEPAKFKMKYWGVASFLQGNDDHWIVDTDYDTYAVQYSCF  
LNLDGTCADSYSFVFSRDPNGLPPEAQKIVRQRQEELCLARQYRLIVHNGYCDGRSERLL
```



Send to: ▾

Choose Destination

- File Clipboard
 Collections Analysis Tool

Download 1 items.

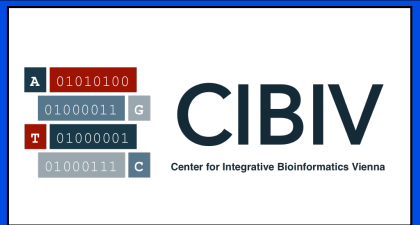
Format

FASTA

Create File

Protein 3D Structu

FASTA format



A sequence in FASTA format

- begins with ">" and a single-line description,
- followed by lines of sequence data.

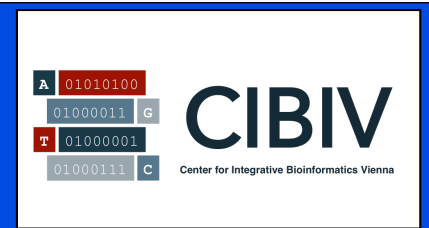
It is recommended that all lines of text be shorter than 80 characters in length.

An example:

```
>P02753
```

```
MKWVWALLLLAALGSGRAERDCRVSSFRVKENFDKARFSGTWYAMAKKDP  
EGLFLQDNIVAEFSVDETGQMSATAKGRVRLNNDVVCADMVGTFTDTE  
PAKFKMKYWGVASFLQKGNDDHWIVDTDYDTYAVQYSCRLNLDGTCADS  
YSFVFSRDPNGLPPEAQKIVRQRQEELCLARQYRLIVHNGYCDGRSERNL  
L
```

Gene entry (summary)



<http://www.ncbi.nlm.nih.gov/gene/5950>

Gene [Save search](#) [Limits](#) [Advanced](#) [Help](#)

Display Settings: Full Report

Send to:

RBP4 retinol binding protein 4, plasma [*Homo sapiens* (human)]

Gene ID: 5950, updated on 15-Sep-2013

Summary

Official Symbol RBP4 provided by HGNC
Official Full Name retinol binding protein 4, plasma provided by HGNC
Primary source [HGNC:9922](#)
Locus tag PRO2222
See related [Ensembl:ENSG00000138207](#); [HPRD:01580](#); [MIM:180250](#); [Vega:OTTHUMG00000018773](#)
Gene type protein coding
RefSeq status REVIEWED
Organism [Homo sapiens](#)
Lineage Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini; Catarrhini; Hominidae; Homo
Also known as RDCCAS
Summary This protein belongs to the lipocalin family and is the specific carrier for retinol (vitamin A alcohol) in the blood. It delivers retinol from the liver stores to the peripheral tissues. In plasma, the RBP-retinol complex interacts with transthyretin which prevents its loss by filtration through the kidney glomeruli. A deficiency of vitamin A blocks secretion of the binding protein posttranslationally and results in defective delivery and supply to the epidermal cells. [provided by RefSeq, Jul 2008]



Table of contents

- Summary
- Genomic context
- Genomic regions, transcripts, and products
- Bibliography
- Phenotypes
- Variation
- Interactions
- Pathways
- General gene information
 - Gene Ontology
- General protein information
- Reference sequences
- Related sequences
- Additional links
 - Locus-specific Databases

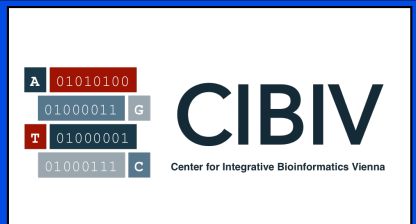
Related information

- Order cDNA clone
- 3D structures
- RinAssay

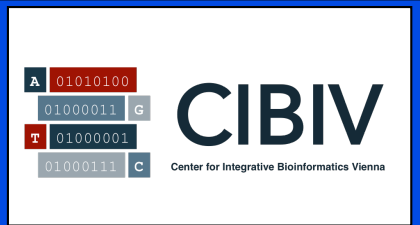


Note, links to many other RBP4 database entries are available

Gene entry (links from genomic region)



Gene entry (links to related sequences)



mRNA and Protein(s)

1. [NM_001323517.1](#) → [NP_001310446.1](#) **retinol-binding protein 4 isoform a precursor**

Status: REVIEWED

| | | |
|---|---|--|
| Description | Transcript Variant: This variant (2) and variant 1 both encode isoform a. | |
| Source sequence(s) | AL356214 | |
| Consensus CDS | CCDS31249.1 | |
| UniProtKB/Swiss-Prot | P02753 | |
| Related | ENSP00000360522.1 , ENST00000371467.5 | |
| Conserved Domains (1) summary | | |
| | pfam00061 Location:39 → 177 | Lipocalin; Lipocalin / cytosolic fatty-acid binding protein family |

2. [NM_001323518.1](#) → [NP_001310447.1](#) **retinol-binding protein 4 isoform b** 

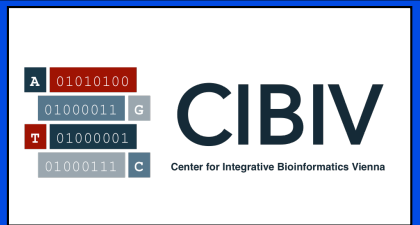
Status: REVIEWED

| | | |
|---|---|--|
| Source sequence(s) | AL356214 , BC020633 , BG565176 , BI712834 | |
| Consensus CDS | CCDS81488.1 | |
| UniProtKB/Swiss-Prot | P02753 | |
| UniProtKB/TrEMBL | Q5VY30 | |
| Related | ENSP00000360524.2 , ENST00000371469.2 | |
| Conserved Domains (1) summary | | |
| | pfam00061 Location:37 → 175 | Lipocalin; Lipocalin / cytosolic fatty-acid binding protein family |

3. [NM_006744.4](#) → [NP_006735.2](#) **retinol-binding protein 4 isoform a precursor**





























See identical proteins and their annotated locations for [NP_006735.2](#)

Gene entry (there are links in all parts)



RBP4 retinol binding protein 4 [*Homo sapiens* (human)]

Gene ID: 5950, updated on 28-May-2019

- Summary  
- Genomic context  
- Genomic regions, transcripts, and products  
- Expression  
- Bibliography  
- Phenotypes  
- Variation  
- Pathways from BioSystems  
- Interactions  
- General gene information  
- General protein information  
- NCBI Reference Sequences (RefSeq)  
- Related sequences  
- Additional links  

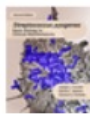
Access to Biomedical Literature



Bookshelf

Bookshelf provides free online access to books and documents in life science and healthcare. Search, read, and discover.

New & Updated



Streptococcus pyogenes: Basic Biology to Clinical Manifestations [Internet]. 2nd edition. Ferretti JJ, Stevens DL, Fischetti VA, editors. Oklahoma City (OK): University of Oklahoma Health Sciences Center; 2022 Oct 8.



Directrices para el diagnóstico y el manejo de la histoplasmosis diseminada en las personas con infección por el VIH [Internet]. Washington (DC): Organización Panamericana de la Salud; 2020 Apr.



Global Roadmap for Healthy Longevity. National Academy of Medicine; Commission for a Global Roadmap for Healthy Longevity. Washington (DC): National Academies Press (US); 2022 Jun 3.



Guidelines for Diagnosing and Managing Disseminated Histoplasmosis among People Living with HIV [Internet]. Washington (DC): Pan American Health Organization; 2020 Apr.



Innovations for Tackling Tuberculosis in the Time of COVID-19: Proceedings of a Workshop. National Academies of Sciences, Engineering, and Medicine; Health and Medicine Division; Board on Global Health; Forum on Microbial Threats; Nicholson A, Liao J, Biffi C, editors. Washington (DC): National Academies Press (US); 2022 Mar 24

Featured Titles



Diagnosis and Management of HIV-2 in Adults [Internet]. Shah SS, Fine SM, Vail RM, et al. Baltimore (MD): Johns Hopkins University; 2022 Aug.



Medical Genetics Summaries [Internet]. Pratt VM, Scott SA, Pirmohamed M, et al., editors. Bethesda (MD): National Center for Biotechnology Information (US); 2012-.



Oral morphine analgesia for preventing pain during invasive procedures in non-ventilated premature infants in hospital: the Poppi RCT. Efficacy and Mechanism Evaluation, No. 6.9. Monk V, Moultrie F, Hartley C, et al. Southampton (UK): NIHR Journals Library; 2019 Aug.



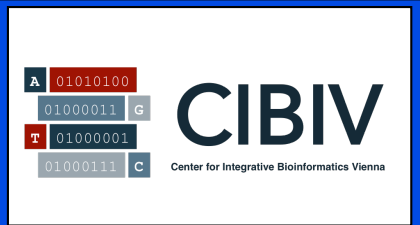
Toward a Post-Pandemic World: Lessons from COVID-19 for Now and the Future: Proceedings of a Workshop. National Academies of Sciences, Engineering, and Medicine; Health and Medicine Division; Board on Global Health; Forum on Microbial Threats; Snair M, Biffi C, Ashby E, editors. Washington (DC): National Academies Press (US); 2022 May 11.



Unequal Treatment: Confronting Racial and Ethnic Disparities in Health Care. Institute of Medicine (US) Committee on Understanding and Eliminating Racial and

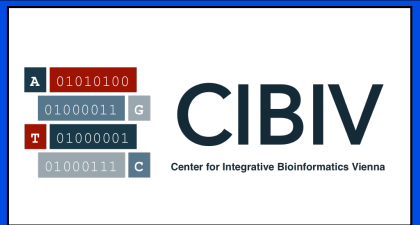
**NCBI
Bookshelf
contains many
online books
that are freely
available**

PubMed to find literature information



- PubMed is the NCBI gateway to MEDLINE.
- MEDLINE contains bibliographic citations
- and author abstracts from over 5,000 journals
- published in the United States and in 70 foreign
- countries.
- It has over 34 million records dating back to the 1960ies.

MeSH indexing in PubMed



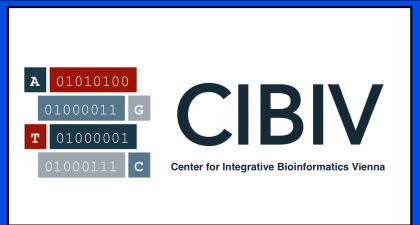
MeSH is the acronym for "Medical Subject Headings."

MeSH is the list of the **vocabulary terms** used for subject analysis of biomedical literature at NLM.

MeSH vocabulary is used for **indexing journal articles** for MEDLINE.

The MeSH controlled vocabulary imposes **uniformity and consistency** to the indexing of biomedical literature.

PubMed search



retinol binding protein fish



Search

[Advanced](#) [Create alert](#) [Create RSS](#)

[User Guide](#)

Save

Email

Send to

Sorted by: Best match

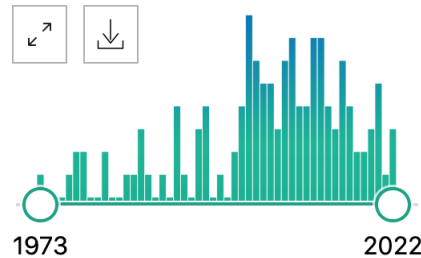
Display options

MY NCBI FILTERS

119 results

Page 1 of 12

RESULTS BY YEAR



TEXT AVAILABILITY

- Abstract
- Free full text
- Full text

ARTICLE ATTRIBUTE

- Associated data

ARTICLE TYPE

- 1 **Carotenoid and retinoid transport to fish oocytes and eggs: what is the role of retinol binding protein?**

Cite Lubzens E, Lissauer L, Levavi-Sivan B, Avarre JC, Sammar M.
Mol Aspects Med. 2003 Dec;24(6):441-57. doi: 10.1016/s0098-2997(03)00040-2.
Share PMID: 14585315 [Review](#).

Fish eggs contain carotenoids, retinals (retinal and dehydroretinal) and retinols (**retinol**, dehydroretinol and retinyl-esters) that are utilized during embryonic development, after fertilization. ...However, the transport of retinols and retinyl-esters that were loc ...

- 2 **Structural and mutational analyses of protein-protein interactions between transthyretin and retinol-binding protein.**

Cite Zanotti G, Folli C, Cendron L, Alfieri B, Nishida SK, Gliubich F, Pasquato N, Negro A, Berni R.
FEBS J. 2008 Dec;275(23):5841-54. doi: 10.1111/j.1742-4658.2008.06705.x.
Share PMID: 19021760 [Free article](#).

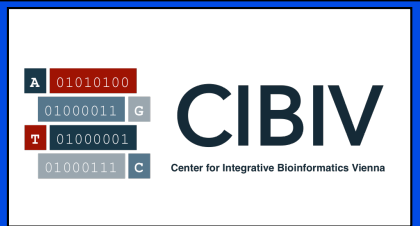
Transthyretin is a tetrameric **binding protein** involved in the transport of thyroid hormones and in the cotransport of **retinol** by forming a complex in plasma with **retinol-binding protein**. ...Remarkably, some of the residues in mutated huma ...

- Hepatic synthesis, maturation and complex formation between retinol-binding**



Back to Top

PubMed search (refinement)



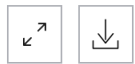
PubMed.gov Sorted by: Best match

MY NCBI FILTERS

119 results

Page 1 of 12

RESULTS BY YEAR



1973 2022

TEXT AVAILABILITY

- Abstract
- Free full text
- Full text

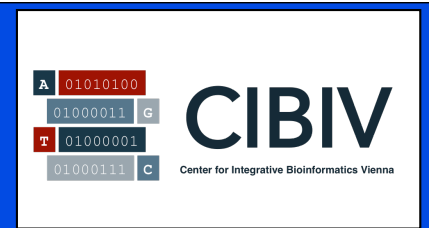
ARTICLE ATTRIBUTE

- Associated data

ARTICLE TYPE

- 1 **Carotenoid and retinoid transport to fish oocytes and eggs: what is the role of retinol binding protein?**
Lubzens E, Lissauer L, Levavi-Sivan B, Avarre JC, Sammar M.
Mol Aspects Med. 2003 Dec;24(6):441-57. doi: 10.1016/s0098-2997(03)00040-2.
PMID: 14585315 [Review](#).
Fish eggs contain carotenoids, retinals (retinal and dehydroretinal) and retinols (**retinol**, dehydroretinol and retinyl-esters) that are utilized during embryonic development, after fertilization. ...However, the transport of retinols and retinyl-esters that were loc ...
- 2 **Structural and mutational analyses of protein-protein interactions between transthyretin and retinol-binding protein.**
Zanotti G, Folli C, Cendron L, Alfieri B, Nishida SK, Gliubich F, Pasquato N, Negro A, Berni R.
FEBS J. 2008 Dec;275(23):5841-54. doi: 10.1111/j.1742-4658.2008.06705.x.
PMID: 19021760 [Free article](#).
Transthyretin is a tetrameric **binding protein** involved in the transport of thyroid hormones and in the cotransport of **retinol** by forming a complex in plasma with **retinol-binding protein**. ...Remarkably, some of the residues in mutated huma ...
- Hepatic synthesis, maturation and complex formation between retinol-binding**

Back to Top



Review > Mol Aspects Med. 2003 Dec;24(6):441-57. doi: 10.1016/s0098-2997(03)00040-2.

Carotenoid and retinoid transport to fish oocytes and eggs: what is the role of retinol binding protein?

E Lubzens¹, L Lissauer, B Levavi-Sivan, J-C Avarre, M Sammar

Affiliations + expand

PMID: 14585315 DOI: 10.1016/s0098-2997(03)00040-2

Abstract

Fish eggs contain carotenoids, retinals (retinal and dehydroretinal) and retinols (retinol, dehydroretinol and retinyl-esters) that are utilized during embryonic development, after fertilization. The carotenoids (mainly astaxanthins) are transported in the plasma by the low density lipoproteins, high density lipoproteins, and very high density lipoproteins (VHDL) and were found to be associated also with serum albumin. Retinals were found to be associated vitellogenin (VTG), a component of the plasma VHDL fraction that is internalized by oocytes during vitellogenesis. However, the transport of retinols and retinyl-esters that were located in the oil droplet fraction of homogenized eggs, has yet to be elucidated. Retinols are more abundant in freshwater fish eggs than in eggs of marine fish species. Since retinol is transported in the plasma of vertebrates in association with retinol binding protein (RBP), recent studies on the molecular characterization and expression sites of RBP, could contribute to determining the involvement of RBP in transporting retinol to developing oocytes in vertebrates. Recently, results from our laboratory show that RBP mRNA levels in the liver and RBP plasma levels did not significantly change with the onset and during vitellogenesis in the Rainbow trout. These results were in contrast with a dramatic elevation in the mRNA levels of VTG in the liver and an increase in VTG plasma levels that was observed in the same females. Moreover, 17beta-estradiol treatment of

FULL TEXT LINKS



ACTIONS

“ Cite

📖 Collections

SHARE



PAGE NAVIGATION

< Title & authors

Abstract

Similar articles

Cited by

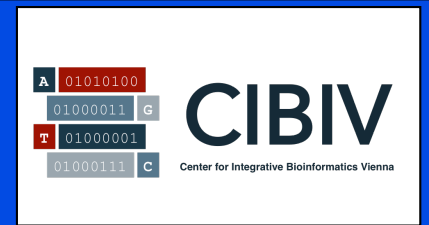
Publication types

NEXT RESULT
2 of 119 >

MeSH terms

At bottom: links to similar articles, citing articles, cross-links, etc

PubMed search strategies



Try the tutorials ("[User Guide](#)" at top)

Use boolean queries (capitalize AND, OR, NOT)
lipocalin AND disease

Utilize search fields (author [au], 1st author [1au], date of publication [dp], title [ti], abstract [ab], title+abstract [tiab], etc):

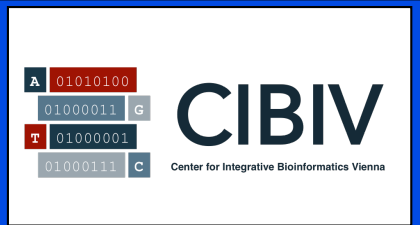
retinol binding protein[TIAB] AND chen[1AU]

Try using "[Advanced](#)"

Try "[Similar articles](#)" for similar articles or "[LinkOut](#)" to other information and external resources

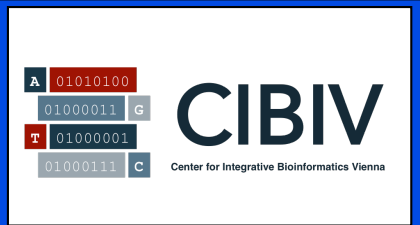
Tools for Sequence Analysis

Tools for Sequence Analysis



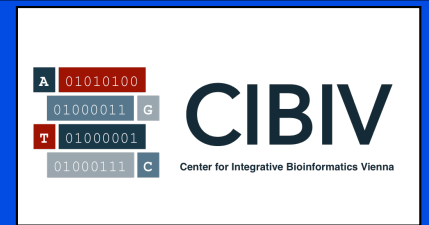
BLAST: The Basic Local Alignment Search Tool (BLAST) for comparing gene and protein sequences against others in public databases, comes in several types including PSI-BLAST, PHI-BLAST, and BLAST 2 sequences.

Tools for Sequence Analysis



[Conserved Domain Database \(CDD\)](#) A collection of sequence alignments and profiles representing protein domains conserved in molecular evolution. The [CD Search Service](#) can be used to search CDD.

Tools for Sequence Analysis

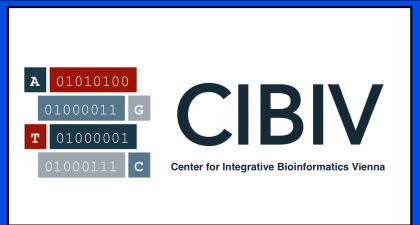


[NCBI Gene](#) Find information on sequence analyses for a particular gene and organism.

[NCBI Protein](#) Same, but protein-centered.

[NCBI Genome](#) Find information about genomes and genome projects.

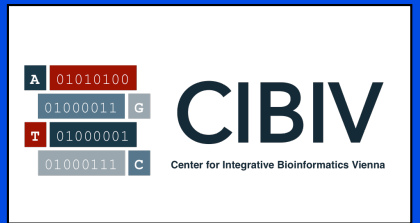
Tools for Sequence Analysis



Gene Expression Omnibus (GEO) GEO provides several tools to assist with the visualization and exploration of curated GEO data.

ORF Finder A graphical analysis tool that finds all open reading frames of a selected minimum size in a user's sequence or in a sequence already in the database.

Tools for Sequence Analysis

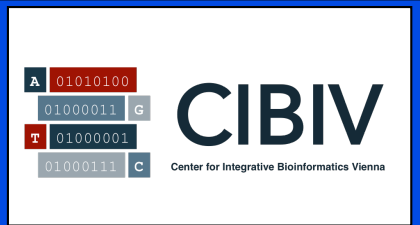


[Trace Archive](#) Developed to store the raw sequence data underlying sequences generated by various genome projects.

[Sequence Read Archive](#) SRA store large amounts of short read data of next-generation sequencing runs generated by various genome projects.

[VecScreen](#) A tool for identifying segments of a nucleic acid sequence that may be of vector, linker, or adapter origin before using Tools for Sequence Analysis or submission.

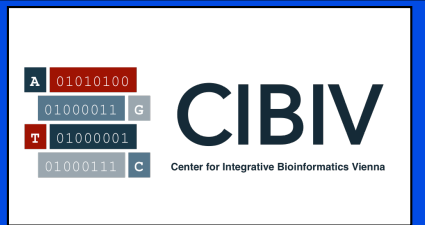
Pairwise Sequence alignment



What is an alignment?

Alignment is the procedure of writing two (or more) protein or DNA sequences in a way that a **maximum of identical or similar characters** are placed in the same column by adding gap characters ('-').

Pairwise Sequence Alignment



What is an alignment?

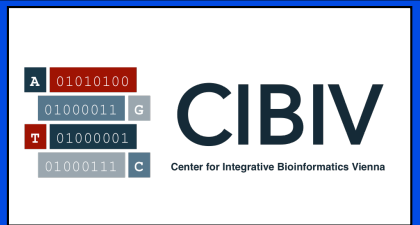
Alignment is the procedure of writing two (or more) protein or DNA sequences in a way that a maximum of identical or similar characters are placed in the same column by adding gap characters ('-').

unaligned sequences:

seq1 : LGPSKQTGASKGSSRIWDN

seq2 : LNTKSAGASKGAILMRLGDAS

Pairwise Sequence Alignment



unaligned sequences:

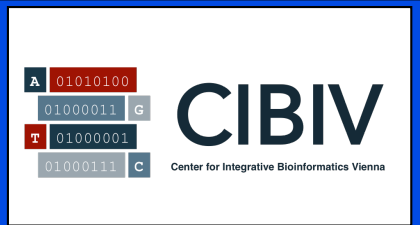
seq1: LGPSKQTGASKGSSRIWDN

seq2: LNTKSAGASKGAILMRLGDAS

aligned sequences:

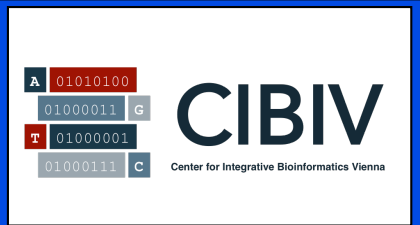
```
seq1:  LGPSKQTGASKGS--SR IWDN-
      |  |  |||||  |  |
seq2:  LN-TKSAGASKGAILMRLGDAS
```

Applications



- Sequence comparison
- Sequencing: to combined sequenced fragments
- Search for genes
- Estimation of evolutionary distance
- Finding genes
- Finding relatives in databases
- Estimating function of genes and proteins
- Estimating structure of RNAs and proteins
- the basis to reconstruct evolutionary relationships and trees

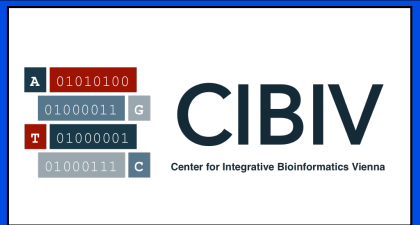
Model



For alignments we consider the following point mutations in comparison of two sequences:

- **substitutions** (change of a character, in alignment: mismatch)
- **insertion** of character(s) in one sequence
- **deletion** of character(s) from one sequence
- identical characters in both sequences are called a **match**

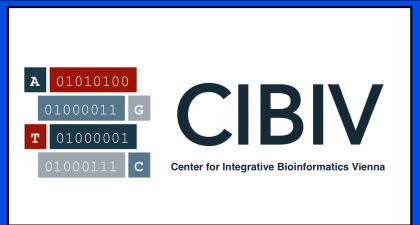
Model



Alignments of two sequences are performed by the the following methods:

- dot matrix analysis or dot plots
- dynamic programming
- word-based or k-tuple methods (BLAST)

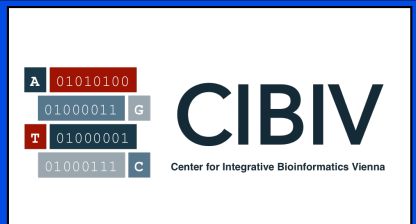
Dot Plot



- Given two sequences $a = \text{GASKGS}$ and $b = \text{GACKGS}$ of length $M = 6$ and $N = 6$.

| | G | A | S | K | G | S |
|---|---|---|---|---|---|---|
| G | | | | | | |
| A | | | | | | |
| C | | | | | | |
| K | | | | | | |
| G | | | | | | |
| S | | | | | | |

Dot Plot



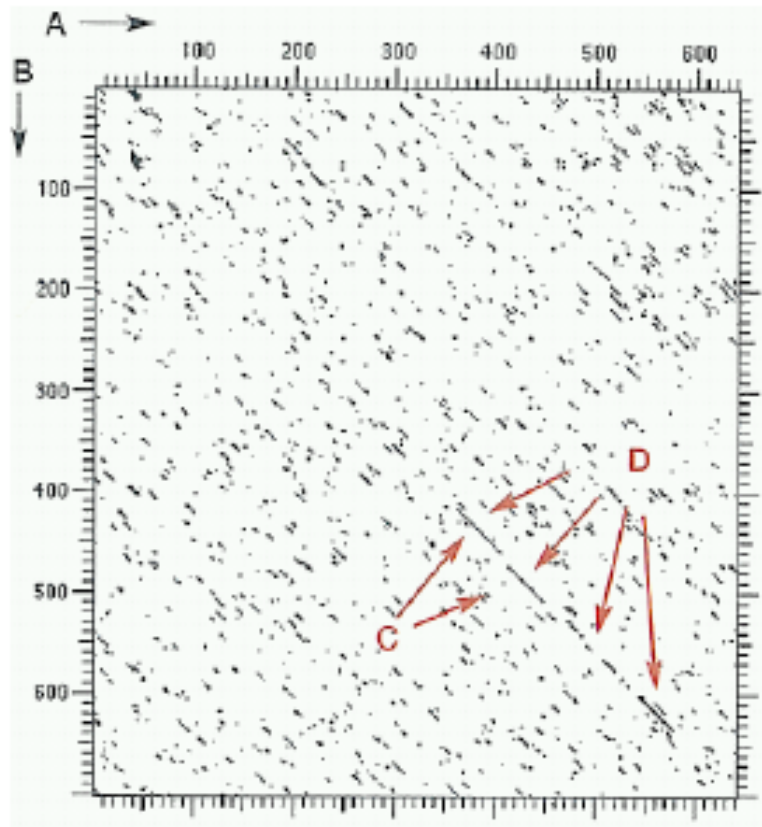
- Given two sequences $a = \text{GASKGS}$ and $b = \text{GACKGS}$ of length $M = 6$ and $N = 6$.

| | G | A | S | K | G | S |
|---|---|---|---|---|---|---|
| G | | | | | | |
| A | | | | | | |
| C | | | | | | |
| K | | | | | | |
| G | | | | | | |
| S | | | | | | |



| | G | A | S | K | G | S |
|---|---|---|---|---|---|---|
| G | ● | | | | ● | |
| A | | ● | | | | |
| C | | | | | | |
| K | | | | ● | | |
| G | ● | | | | ● | |
| S | | | ● | | | ● |

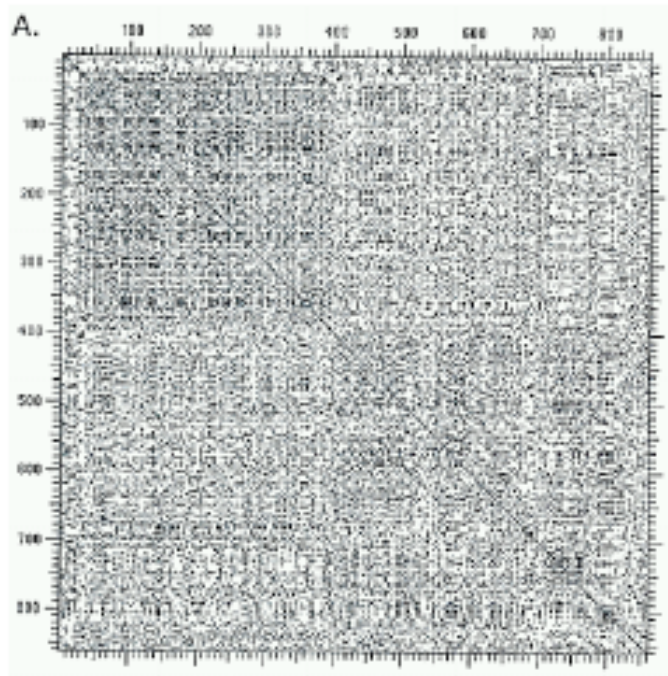
Dot Plot: What can we learn



- existing alignable sequences
- possible indels
- duplicated sequences and repeats
- self-complementarity
- gene-orders between genomes

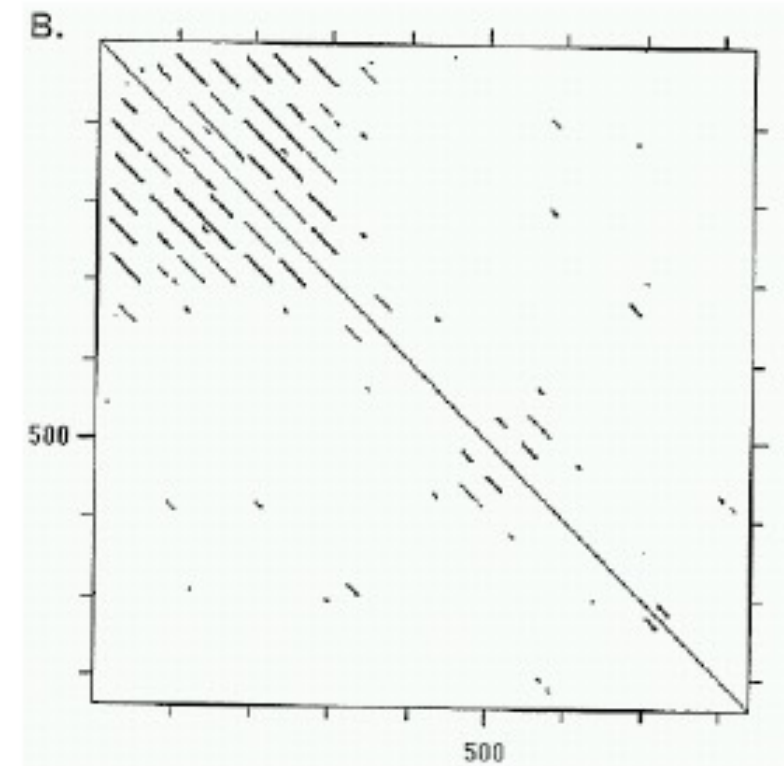
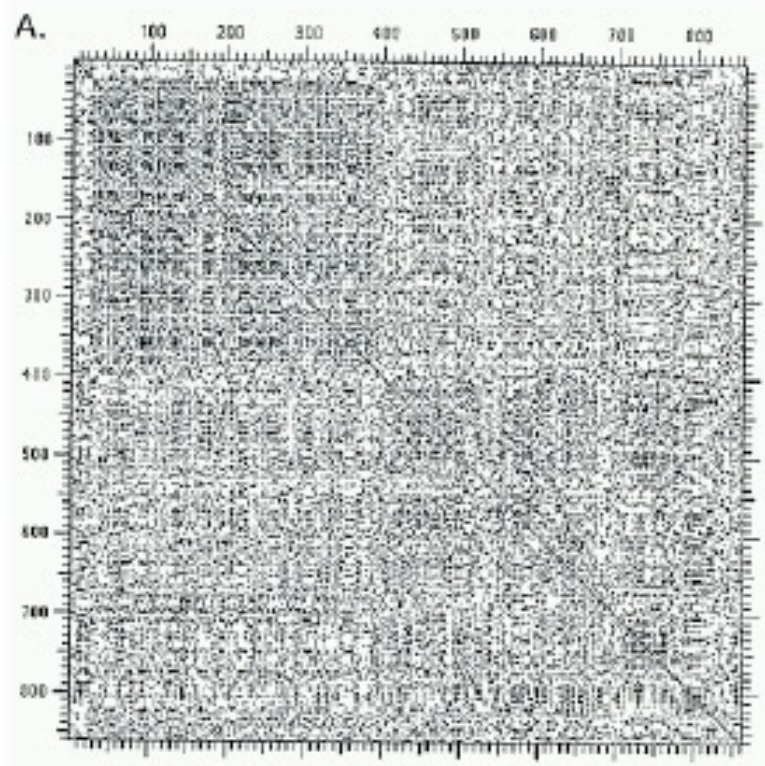
Dot Plot: Filtering

Especially with DNA sequences, dot plots may show a lot of noise. Filtering should be applied to improve the result:



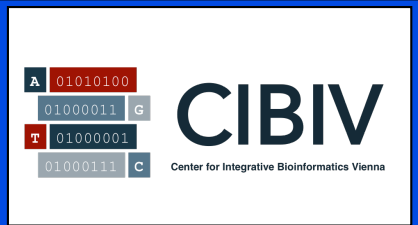
- Draw a dot only if there are more than v matches (stringency) on a diagonal (window) of length w .
- Typical values for DNA sequences are: window 15 *bp*, stringency of 10 matches.
- Proteins sequences are usually not filtered, due to their high number character states: 20 amino acids compared to 4 bases in DNA.
- For very dissimilar proteins with few matches, filtering with window of 20 and stringency 5 might visualize any similarity.

Dot Plot: Example



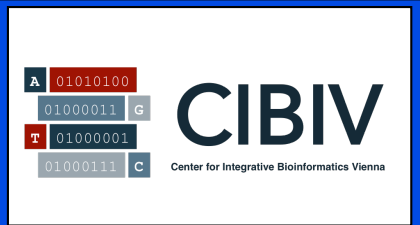
Filtering with windows of 23, stringency 7 makes the repetitive sequence visible.

Dot Plot: Drawbacks



- A dot plot does not give us an actual alignment,
- it only visualized possible alignable regions.
- A dot plot need to be visible on the screen, which is impossible for long sequences, getting uncomprehensible or showing too little resolution.

Dynamic Programming

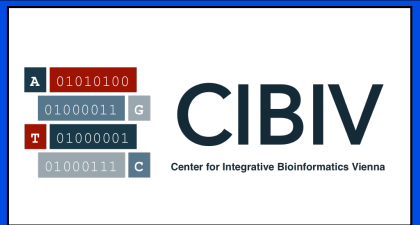


One way to get an alignment, is the application of the **dynamic programming** principle by *Bellman* from the early 1950ies.

A **dynamic programming** approach usually incorporates

- a recursive mathematical description of the optimal solution.
- the computation of all intermediate values needed to find the optimal solution (by avoiding double-computations).
- The *bottom-up* construction of the optimal solution from the above values.

Dynamic Programming: The Basics



Given two sequences A and B and a scoring function for two characters a and b

$$S(a, b) = \begin{cases} +5 & \text{if } a = b \text{ (match)} \\ -2 & \text{if } a \neq b \text{ (mismatch)} \\ -6 & \text{if } a \text{ or } b \text{ indel (gap)} \end{cases}$$

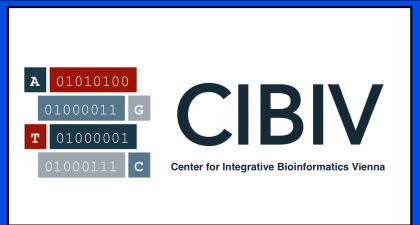
to score each alignment column.

Then we are looking for that alignment, that gives us the highest score $S(A, B)$ summing up the column scores $s(a, b)$ for all columns of the alignment.

For example:

$$\begin{array}{cccccccc} \mathbf{T} & \mathbf{G} & \mathbf{C} & \mathbf{T} & \mathbf{C} & \mathbf{G} & \mathbf{T} & \mathbf{A} \\ \mathbf{T} & \mathbf{-} & \mathbf{-} & \mathbf{T} & \mathbf{C} & \mathbf{A} & \mathbf{T} & \mathbf{A} \\ +5 & -6 & -6 & +5 & +5 & -2 & +5 & +5 = 11 \end{array}$$

Dynamic Programming: The Basics 2



- There are be far too many.
- There are about $\frac{2^{2N}}{\sqrt{2\pi N}}$ possible alignments,
- for two sequences of length $N = 300$ that is 10^{179} alignments.

Hence, we need a smart way to cut the computation short, like the **dynamic programming** approach for pairwise alignment by *Needleman and Wunsch* (1970).

Needleman Wunsch: Global Alignment

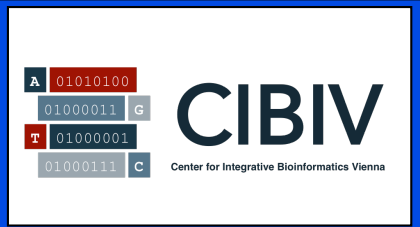


Given sequences A and B and scoring function $s(a, b) = \begin{cases} +5 & a = b \\ -2 & a \neq b \\ -6 & a \text{ or } b \text{ indel} \end{cases}$

| | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---|-----|-----|-----|-----|-----|-----|-----|-----|-----------|
| | T | G | A | A | C | G | T | A | |
| 0 | 0 | -6 | -12 | -18 | -24 | -30 | -36 | -42 | -48 |
| 1 | -6 | 5 | -1 | -7 | -13 | -19 | -25 | -31 | -37 |
| 2 | -12 | -1 | 3 | 4 | -2 | -8 | -14 | -20 | -26 |
| 3 | -18 | -7 | -3 | 1 | 2 | 3 | -3 | -9 | -15 |
| 4 | -24 | -13 | -9 | 2 | 6 | 0 | 1 | -5 | -4 |
| 5 | -30 | -19 | -15 | -4 | 0 | 4 | -2 | 6 | 0 |
| 6 | -36 | -25 | -21 | -10 | 1 | -2 | 2 | 0 | 11 |

- Initialize an $N \times M$ matrix with the sequences A and B of length M and N .
- Starting at the upper left corner set the intermediate scoring value $\sigma(i, j) = \max \begin{cases} \sigma(i-1, j-1) + s(A_i, B_j) & \text{match/mismatch} \\ \sigma(i-1, j) + s(A_i, -) & \text{gap in } B \\ \sigma(i, j-1) + s(B_i, -) & \text{gap in } A \end{cases}$
- $\sigma(i, j)$ always holds the optimal score for the alignment from the sequence start to (A_i, B_j) .
- The optimal score can be found at $\sigma(N, M)$.

Needleman Wunsch: Global Alignment



Given sequences A and B and scoring function $s(a, b) = \begin{cases} +5 & a = b \\ -2 & a \neq b \\ -6 & a \text{ or } b \text{ indel} \end{cases}$

| | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| | T | G | A | A | C | G | T | A | |
| 0 | 0 | -6 | -12 | -18 | -24 | -30 | -36 | -42 | -48 |
| 1 | -6 | 5 | -1 | -7 | -13 | -19 | -25 | -31 | -37 |
| 2 | -12 | -1 | 3 | 4 | -2 | -8 | -14 | -20 | -26 |
| 3 | -18 | -7 | -3 | 1 | 2 | 3 | -3 | -9 | -15 |
| 4 | -24 | -13 | -9 | 2 | 6 | 0 | 1 | -5 | -4 |
| 5 | -30 | -19 | -15 | -4 | 0 | 4 | -2 | 6 | 0 |
| 6 | -36 | -25 | -21 | -10 | 1 | -2 | 2 | 0 | 11 |

Resulting alignment and score:

T G A A C G T A

T - - A C A T A

$$+5 - 6 - 6 + 5 + 5 - 2 + 5 + 5 = 11$$

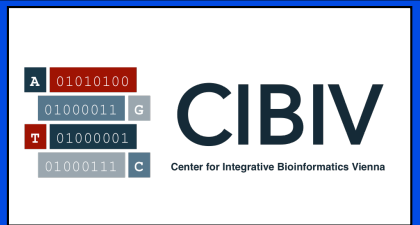
T G A A C G T A

T - A - C A T A

$$+5 - 6 + 5 - 6 + 5 - 2 + 5 + 5 = 11$$

- Initialize an $N \times M$ matrix with the sequences A and B of length M and N .
- Starting at the upper left corner set the intermediate scoring value $\sigma(i, j) = \max \begin{cases} \sigma(i-1, j-1) + s(A_i, B_j) & \text{match/mismatch} \\ \sigma(i-1, j) + s(A_i, -) & \text{gap in } B \\ \sigma(i, j-1) + s(B_i, -) & \text{gap in } A \end{cases}$
- $\sigma(i, j)$ always holds the optimal score for the alignment from the sequence start to (A_i, B_j) .
- The optimal score can be found at $\sigma(N, M)$.
- The optimal alignment is retrieved by following the best values $\sigma(i, j)$.
- There can be more than one optimal alignment!

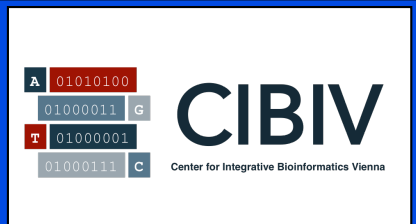
Smith Waterman: Local Alignment



Temple Smith and Mike Waterman (1981) found a way to adapt the Needleman-Wunsch algorithm to produce local alignments:

- the scoring function must contain negative values for mismatches,
- whenever $\sigma(i, j) > 0$ it is set to zero (here a possible backtrace will stop)
- the initial row and column with gap costs are set to zero
- the backtrace will start at the maximal $\sigma(i, j)$

Smith Waterman: Local Alignment



Given sequences A and B and scoring function $s(a, b) = \begin{cases} +5 & a = b \\ -2 & a \neq b \\ -6 & a \text{ or } b \text{ indel} \end{cases}$

| | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|----|---|----|----|
| | T | G | C | T | C | G | T | G | |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 1 | 0 | 5 | 0 | 0 | 5 | 0 | 0 | 5 | 0 |
| 2 | 0 | 5 | 3 | 0 | 5 | 3 | 0 | 5 | 3 |
| 3 | 0 | 0 | 3 | 8 | 2 | 10 | 4 | 0 | 3 |
| 4 | 0 | 0 | 0 | 2 | 6 | 4 | 8 | 2 | 0 |
| 5 | 0 | 5 | 0 | 0 | 7 | 4 | 2 | 13 | 7 |
| 6 | 0 | 0 | 3 | 0 | 1 | 5 | 2 | 7 | 11 |

Resulting alignment and score:

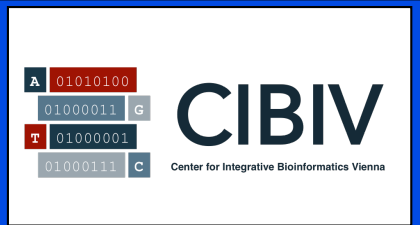
T C G T

T C A T

$$+5+5-2+5=13$$

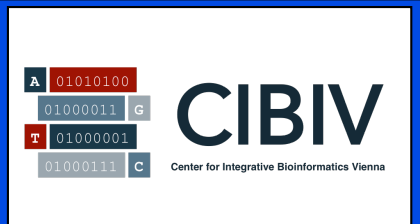
- Initialize an $N \times M$ matrix with the sequences A and B of length M and N .
- Starting at the upper left corner set the intermediate scoring value $\sigma(i, j) = \max \begin{cases} \sigma(i-1, j-1) + s(A_i, B_j) & \text{match/mismatch} \\ \sigma(i-1, j) + s(A_i, -) & \text{gap in } B \\ \sigma(i, j-1) + s(-, B_j) & \text{gap in } A \\ 0 & \end{cases}$
- The optimal local alignment score is the maximal score among all $\sigma(i, j)$.
- The optimal local alignment is retrieved by backtracking until the $\sigma(i, j)$ of the current cell (i, j) gets zero.

BLAST



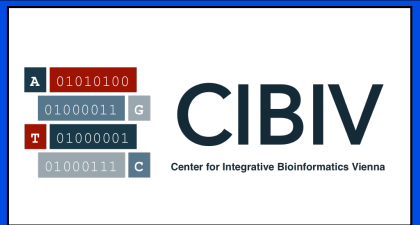
- BLAST (Basic Local Alignment Search Tool; Altschul, Gish, Miller, Myers, Lipman, 1990)
 - later: PsiBLAST, PhiBLAST (Altschul et al. 1997)

BLAST



- Published and released in 1990
- as a public, cost-free service of the NCBI, NIH Washington, DC
- Large server infrastructure to support thousands of requests
- ⇒ the dominant server resource for sequence searches
- Features:
 - Speed (algorithm: sublinear approximate matching; Myers, 1994)
 - outputs range of solutions and statistics (theory: Karlin Altschul, 1990-93)
 - each alignment accompanied with statistical significance (probability that a match of that score or better occurs by chance in aligning random strings)

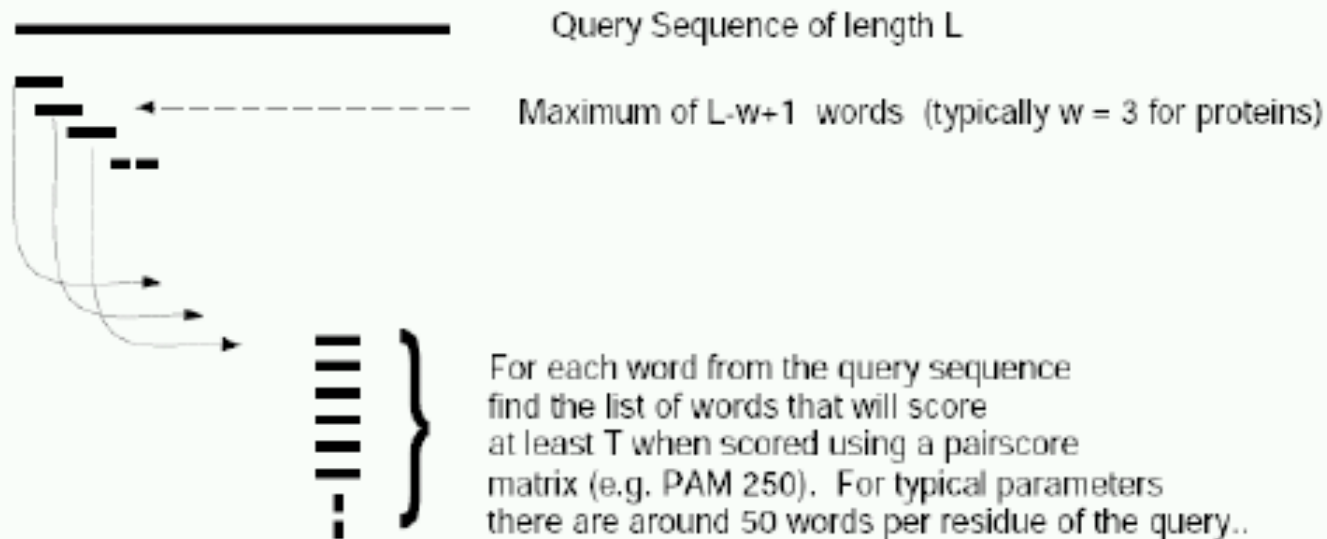
BLAST



- Compile list of high-scoring strings (HSPs, words)
- search for hits → seeds
- extend seeds

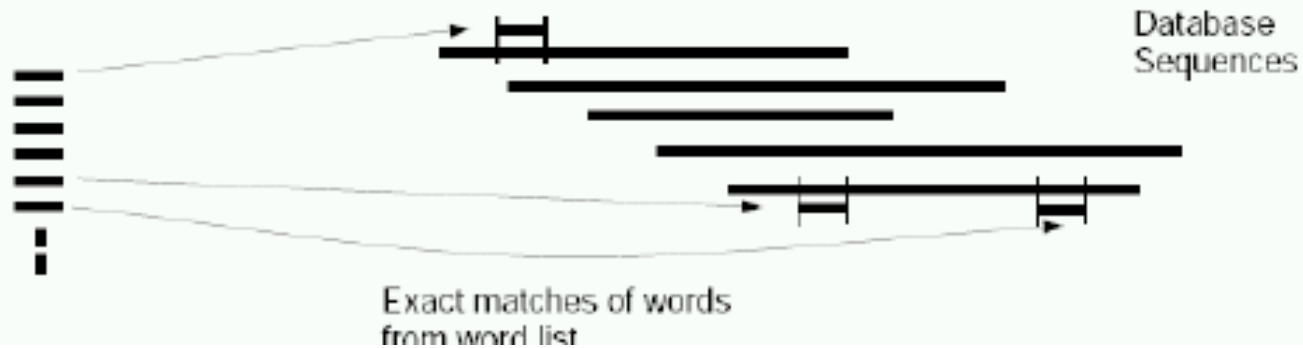
BLAST

(1) For the query find the list of high scoring words of length w .



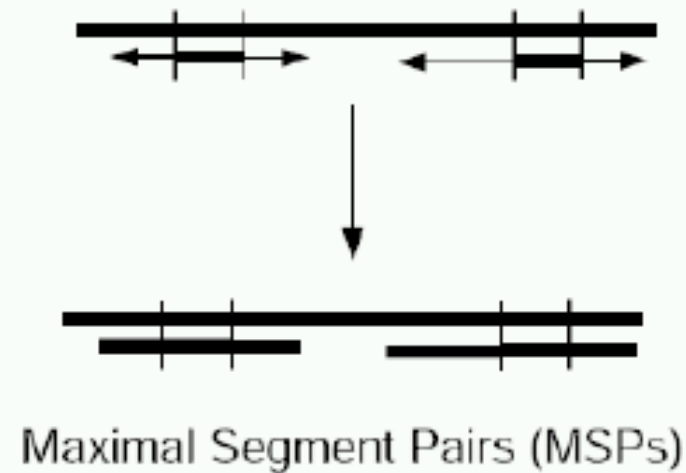
BLAST

(2) Compare the word list to the database and identify exact matches.

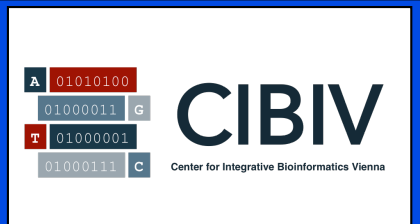


BLAST

- (3)** For each word match, extend alignment in both directions to find alignments that score greater than score threshold S .



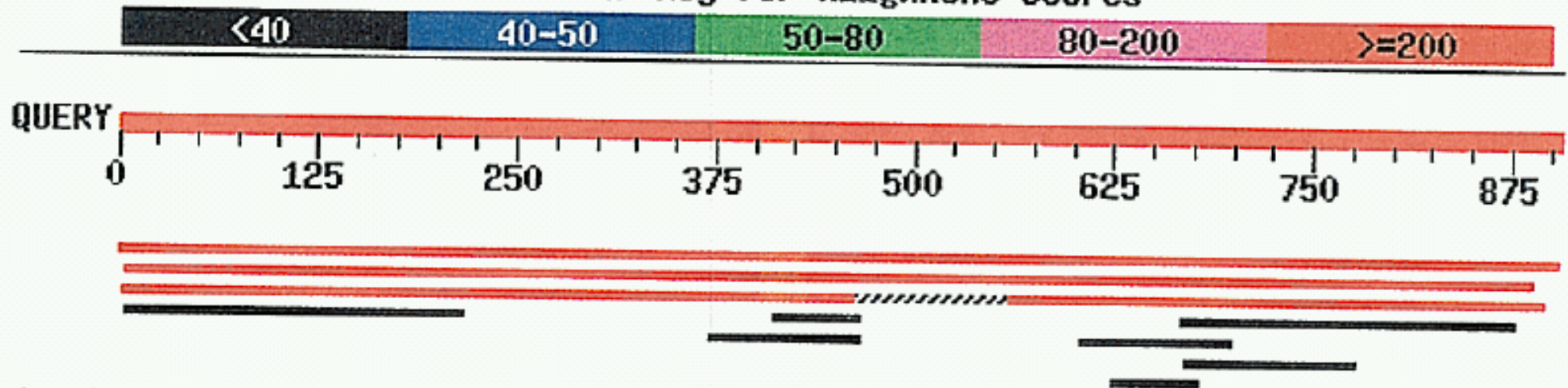
BLAST



A BLASTP 2.0.5 [May-5-1998]
Query= human XP-F repair gene (905 letters)

Database: Non-redundant SwissProt sequences 74,596 sequences; 26,848,718 total letters

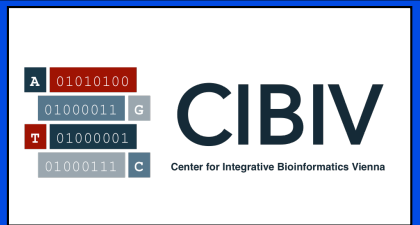
B Color Key for Alignment Scores



Distribution of 11 BLAST Hits on the Query Sequence

| Sequences producing significant alignments: | | | | Score | E |
|---|------------|--|------|--------|-------|
| | | | | (bits) | Value |
| sp Q92889 | XPF_HUMAN | DNA-REPAIR PROTEIN COMPLEMENTING XP-F CELL ... | 1659 | 0.0 | |
| sp P36617 | RA16_SCHPO | DNA REPAIR PROTEIN RAD16 | 485 | e-136 | |
| sp P06777 | RAD1_YEAST | DNA REPAIR PROTEIN RAD1 | 231 | 4e-60 | |
| sp P40562 | YIS2_YEAST | PUTATIVE ATP-DEPENDENT RNA HELICASE YIR002C | 37 | 0.17 | |
| sp Q10202 | YAXB_SCHPO | PUTATIVE ATP-DEPENDENT RNA HELICASE C13F4.11C | 36 | 0.38 | |

References



Possibly the most comprehensive ones:

- The NCBI Handbook, 2nd Ed.:
<http://www.ncbi.nlm.nih.gov/books/NBK143764/>
- They have chapters about many of the things discussed and much more.