

**Heiko Schmidt**

Center for Integrative Bioinformatics Vienna (CIBIV)  
<http://www.cibiv.at>

## Nucleotide Databases

**Three major public DNA databases**

**EMBL/ENA** ↔ **GenBank** ↔ **DDBJ**

**The underlying raw DNA sequences are identical**

# 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 <b>nucleotide and amino acid</b> sequences
Sequence Read Archive (SRA)	<b>sequencing data</b> from high-throughput sequencing platforms like Illumina, Roche 454, PacBio
Genome	sequence and map data from <b>whole genomes</b> of organisms (573 Archaea, 8573 Bacteria, 1936 Eukaryotes, 5482 Viruses), complete and in progress
HomoloGene	A <b>gene homology</b> database that compares nucleotide sequences between pairs of organisms to identify putative orthologs
Taxonomy	names and <b>phylogenetic lineages</b> of more than 360,000 organisms

# Nucleotide Databases



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

NCBI



**National Center for Biotechnology  
Information (NCBI)**

[www.ncbi.nlm.nih.gov](http://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. 12 Jun 2019
- Presentation on NCBI's genome browser at Rocky Mountain Genomics Hackcon 12 Jun 2019
- On June 18, 2019, NCBI's Sanjida



## PubMed

PubMed comprises more than 29 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.

- |   |   |  |
|---|---|--|
| <h4>Using PubMed</h4> <ul style="list-style-type: none"> <li><a href="#">PubMed Quick Start Guide</a></li> <li><a href="#">Full Text Articles</a></li> <li><a href="#">PubMed FAQs</a></li> <li><a href="#">PubMed Tutorials</a></li> <li><a href="#">New and Noteworthy</a></li> </ul>               | <h4>PubMed Tools</h4> <ul style="list-style-type: none"> <li><a href="#">PubMed Mobile</a></li> <li><a href="#">Single Citation Matcher</a></li> <li><a href="#">Batch Citation Matcher</a></li> <li><a href="#">Clinical Queries</a></li> <li><a href="#">Topic-Specific Queries</a></li> </ul>  | <h4>More Resources</h4> <ul style="list-style-type: none"> <li><a href="#">MeSH Database</a></li> <li><a href="#">Journals in NCBI Databases</a></li> <li><a href="#">Clinical Trials</a></li> <li><a href="#">E-Utilities (API)</a></li> <li><a href="#">LinkOut</a></li> </ul> |
| <h4>Latest Literature</h4> <p>New articles from highly accessed journals</p> <ul style="list-style-type: none"> <li><a href="#">Am J Clin Nutr (2)</a></li> <li><a href="#">Cochrane Database Syst Rev (5)</a></li> <li><a href="#">Drugs (1)</a></li> <li><a href="#">J Biol Chem (5)</a></li> </ul> | <h4>Trending Articles</h4> <p>PubMed records with recent increases in activity</p> <ul style="list-style-type: none"> <li><a href="#">Transposon-encoded CRISPR-Cas systems direct RNA-guided DNA integration. Nature. 2019.</a></li> <li><a href="#">Vitamin D Supplementation and Prevention of Type 2 Diabetes. N Engl J Med. 2019.</a></li> </ul> |  |

# NCBI PubMed



NCBI Resources How To

Sign in to NCBI

PubMed.gov  
US National Library of Medicine  
National Institutes of Health

PubMed



Advanced

Search

Help

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 29 million entries** (reference and abstracts)
- links to participating online journals

# NCBI (Entrez) System



NCBI

National Center for  
Biotechnology Information

All Databases



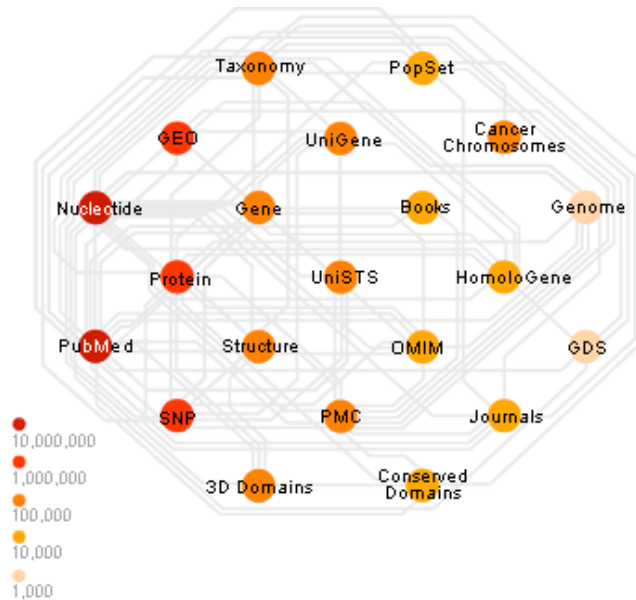
The NCBI/Entrez system integrates...

- 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



Entrez is a search and retrieval system that integrates NCBI databases



# NCBI OMIM



NCBI Resources How To haschmidt My NCBI Sign Out

OMIM OMIM Search Limits Advanced 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](http://omim.org).

Using OMIM	OMIM tools	Related Resources
<a href="#">Getting Started</a>	<a href="#">OMIM API</a>	<a href="#">ClinVar</a>
<a href="#">FAQ</a>		<a href="#">Gene</a>
		<a href="#">GTR</a>
		<a href="#">MedGen</a>

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.

<b>Using Bookshelf</b> <a href="#">Quick Start Guide</a> <a href="#">FAQ</a>	<b>Read</b> <a href="#">Browse Titles</a> <a href="#">New Releases</a>	<b>Participate</b> <a href="#">Authors and Publishers</a> <a href="#">How to Apply</a>
--	--	--

## Bookshelf is...

- searchable resource of on-line books

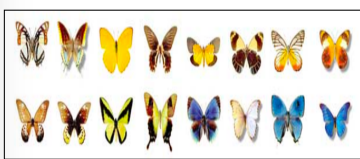
# NCBI Taxonomy



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.

<b>Using Taxonomy</b> <a href="#">Quick Start Guide</a> <a href="#">FAQ</a>	<b>Taxonomy Tools</b> <a href="#">Browser</a> <a href="#">Common Tree</a>	<b>Other Resources</b> <a href="#">GenBank</a> <a href="#">LinkOut</a>
---	---	--

## 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

## 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.

## What is an accession number?

An accession number is **label** that 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      |



### Several ways to access DNA and protein sequences

- [1] **NCBI Gene with RefSeq**  
<https://www.ncbi.nlm.nih.gov/>
- [2] European Bioinformatics Institute (EBI)  
with InterPro and Ensembl  
(separate from NCBI)  
<https://www.ebi.ac.uk/>  
<https://www.ebi.ac.uk/interpro/>  
<http://www.ensembl.org/>

### Several ways to access DNA and protein sequences

- [3] ExPASy, SIB Bioinformatics Resource Portal  
(separate from NCBI, Switzerland)  
<https://www.expasy.org/>
- [4] Universal Protein Resource (UniProt)  
(separate from NCBI)  
<https://www.uniprot.org/>

# NCBI Databases



## 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"

The screenshot shows the NCBI homepage with the search bar containing the text "rbp4". An orange arrow points to the search bar. The page includes a navigation menu on the left, a "Welcome to NCBI" message, and sections for "Popular Resources", "Develop", "Analyze", and "Research".

# NCBI Databases: (2019)



Literature		Genes		Genetics	
Bookshelf	9	Gene	380	ClinVar	29
MeSH	5	GEO DataSets	2,073	dbGaP	0
NLM Catalog	2	GEO Profiles	9,450	dbSNP	6,488
PubMed	977	HomoloGene	3	dbVar	234
PubMed Central	2,034	PopSet	2	GTR	33
		UniGene	17	MedGen	4
				OMIM	11

Proteins		Genomes		Chemicals	
Conserved Domains	0	Assembly	0	BioSystems	418
Identical Protein Groups	14	BioCollections	0	PubChem BioAssay	136
Protein	565	BioProject	2	PubChem Compound	0
Protein Clusters	0	BioSample	2,065	PubChem Substance	122
Sparcle	1	Genome	60		
Structure	44	Nucleotide	1,068		
		Probe	153		
		SRA	3,756		
		Taxonomy	0		

# NCBI Gene



Gene   
Create RSS Create alert 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](#)

GENE Was this helpful?

[RBP4 orthologs from vertebrates](#)

retinol binding protein 4

[How are orthologs calculated?](#)

[Genes with similar protein architectures](#)

[How are similar genes calculated?](#)

Filters: [Manage Filters](#)

**Results by taxon**

Top Organisms [\[Tree\]](#)

- Homo sapiens (50)
- Drosophila melanogaster (26)
- Mus musculus (16)
- Sus scrofa (5)
- Rattus norvegicus (2)
- All other taxa (281)

[More...](#)

**Find related data**

Database:

**Search results**

Items: 1 to 20 of 380 << First < Prev Page 1 of 19 Next > Last >>

Name/Gene ID	Description	Location	Aliases	MIM
<input type="checkbox"/> <a href="#">RBP4</a> ID: 5950	retinol binding protein 4 [Homo sapiens (human)]	Chromosome 10, NC_000010.11 (93591836..93601744, complement)	MCOPCB10, RDCCAS	180250

**Search details**

rbp4[All Fields] AND alive[prop]

[See more...](#)

# NCBI Gene (refine search)



Gene    [Create RSS](#) [Create alert](#) [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: [Manage Filters](#) [Hide sidebar >>](#)

GENE Was this helpful?

[RBP4 orthologs from vertebrates](#)

retinol binding protein 4

[How are orthologs calculated?](#)

[Genes with similar protein architectures](#)

[How are similar genes calculated?](#)

**Search results**  
Items: 1 to 20 of 380 << First < Prev Page 1 of 19 Next > Last >>  
[See also 7 discontinued or replaced items.](#)

Name/Gene ID	Description	Location	Aliases	MIM
<input type="checkbox"/> <a href="#">RBP4</a> ID: 5950	retinol binding protein 4 [ <i>Homo sapiens</i> (human)]	Chromosome 10, NC_000010.11 (93591836..93601744, complement)	MCOPCB10, RDCCAS	180250

**Results by taxon**  
Top Organisms [\[Tree\]](#)  
Homo sapiens (50)  
Drosophila melanogaster (26)  
Mus musculus (16)  
Sus scrofa (5)  
Rattus norvegicus (2)  
All other taxa (281)  
More...

**Find related data**  
Database:

**Search details**  
rbp4[All Fields] AND alive[prop]  
 [See more...](#)

# Filters reduce the number of entries



Gene    [Create RSS](#) [Create alert](#) [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: [Manage Filters](#) [Hide sidebar >>](#)

**Search results**  
Items: 1 to 20 of 50 << First < Prev Page 1 of 3 Next > Last >>  
[Showing Current items.](#)

Name/Gene ID	Description	Location	Aliases	MIM
<input type="checkbox"/> <a href="#">RBP4</a> ID: 5950	retinol binding protein 4 [ <i>Homo sapiens</i> (human)]	Chromosome 10, NC_000010.11 (93591836..93601744, complement)	MCOPCB10, RDCCAS	180250
<input type="checkbox"/> <a href="#">POLR2D</a> ID: 5433	RNA polymerase II subunit D [ <i>Homo sapiens</i> (human)]	Chromosome 2, NC_000002.12 (127846266..127858155, complement)	HSRBP4, HSRPB4, RBP4, RPB16	603000
<input type="checkbox"/> <a href="#">ADIPOQ</a> ID: 9370	adiponectin, C1Q and collagen domain containing [ <i>Homo sapiens</i> (human)]	Chromosome 3, NC_000003.12 (186842674..186858463)	ACDC, ACRP30, ADIPQTL1, ADPN, APM-1, APM1, GBP28	603000
<input type="checkbox"/> <a href="#">LEP</a> ID: 3952	leptin [ <i>Homo sapiens</i> (human)]	Chromosome 7, NC_000007.14 (128241201..128257629)	LEPD, OB, OBS	162950
<input type="checkbox"/> <a href="#">ACE</a> ID: 1636	angiotensin I converting enzyme [ <i>Homo sapiens</i> (human)]	Chromosome 17, NC_000017.11 (63477061..63498380)	ACE1, CD143, DCP, DCP1	103550
<input type="checkbox"/> <a href="#">IL1B</a> ID: 3553	interleukin 1 beta [ <i>Homo sapiens</i> (human)]	Chromosome 2, NC_000002.12 (112829758..112836843,	IL-1, IL1-BETA, IL1F2	146250

**Find related data**  
Database:

**Search details**  
rbp4[All Fields] AND "Homo sapiens"[porgn] AND alive[prop]  
 [See more...](#)

**Recent activity**  
[Turn Off](#) [Clear](#)

- (rbp4) AND "Homo sapiens"[porgn] AND alive[prop] (50) Gene
- (rbp4 AND "Homo sapiens") AND "Homo sapiens"[porgn] AND alive[prop] (51) Gene
- rbp4 AND "Homo sapiens" AND alive[prop] (380) Gene
- entrez (915) Books

# Gene entry (summary)



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

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

### 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](#)
- 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](#)

# Gene entry (genomic region)

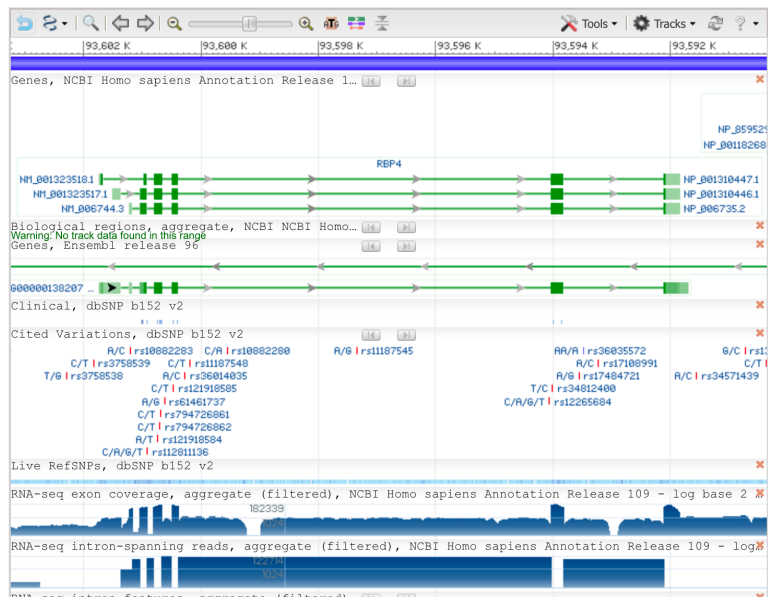


### Genomic regions, transcripts, and products

Go to [reference sequence details](#)

Genomic Sequence:  Chromosome 10 Reference GRCh38.p12 Primary Assembly

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



# Gene entry (expression)



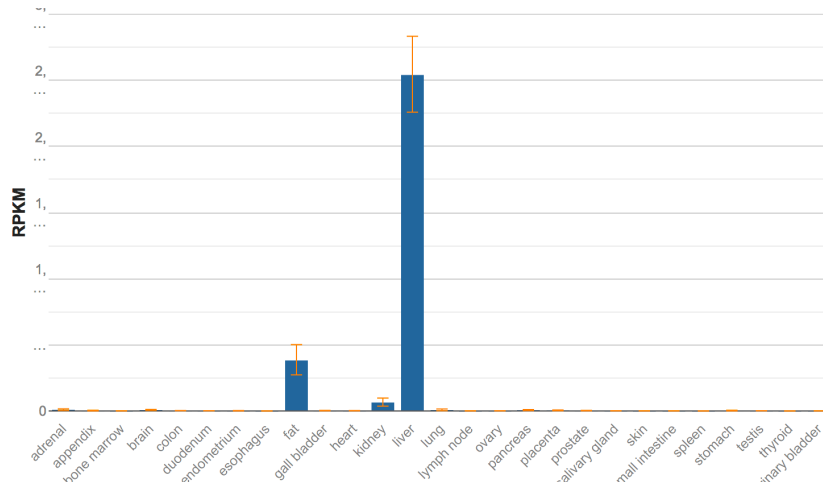
## Expression



HPA RNA-seq normal tissues

[See details](#)

- 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) ]

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

# Gene entry (protein via RefSeq)



## 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)	<a href="#">AL356214</a>	
Consensus CDS	<a href="#">CCDS31249.1</a>	
UniProtKB/Swiss-Prot	<a href="#">P02753</a>	
Related	<a href="#">ENSP00000360522.1</a> , <a href="#">ENST00000371467.5</a>	
Conserved Domains (1) <a href="#">summary</a>	<a href="#">pfam00061</a> Lipocalin; Lipocalin / cytosolic fatty-acid binding protein family Location:39 → 177	

2. [NM\\_001323518.1](#) → [NP\\_001310447.1](#) **retinol-binding protein 4 isoform b**

Status: REVIEWED

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

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)	<a href="#">AL356214</a> , <a href="#">BC020633</a> , <a href="#">BX495987</a> , <a href="#">X00129</a>	
Consensus CDS	<a href="#">CCDS31249.1</a>	
UniProtKB/Swiss-Prot	<a href="#">P02753</a>	
Related	<a href="#">ENSP00000360519.3</a> , <a href="#">ENST00000371464.8</a>	
Conserved Domains (1) <a href="#">summary</a>	<a href="#">pfam00061</a> Lipocalin; Lipocalin / cytosolic fatty-acid binding protein family Location:39 → 177	

# Protein entry (top)



Protein

GenPept ▾

Send to: ▾

Change region shown ▾

## 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

Customize view ▾

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	Location/Qualifiers
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"
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; pfam0061" /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 mkwvwal111 aalgsgrae dcrvssfrvk enfdkarfsg twyamakkdp eglflgdniv 61 aefsvdetgg msatakgrvr llndwvoad mvgtftdted pakfkmkywg vasflqkgn 121 dhwvtdtdyd tyavqysr1 lnlldtcaads ysfvrsrdpn glppeaqkiv rqrqeelcla 181 rgyrlivhng yodgrsernl 1

# Protein entry (top)



Protein

GenPept

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

NCBI Reference Sequence: NP\_001310446.1  
[Identical Proteins](#) [FASTA](#) [Graphics](#)

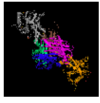
Go to:

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, Araae 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: 2WVQA  
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]
MKVWVALLLLAALGSGRAERDCRVSSFRVKNFVKARFSGTWYAMAKKDPEGLFLQDNIVAEFVSVDETQ
MSATAKGRVRLNNWDVCDMVGTFDTEDPAKFKMKYWGVASFLQKGNDDHWIVD TDYDTYAVQYSCRL
LNL DGT CADSY SFVFSRDPNGLPPEAQKIVRQRQEELCLARQYRLIVHNGYCDGRSERNLL
```

# Protein entry (download, FASTA format)



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]
MKVWVALLLLAALGSGRAERDCRVSSFRVKNFVKARFSGTWYAMAKKDPEGLFLQDNIVAEFVSVDETQ
MSATAKGRVRLNNWDVCDMVGTFDTEDPAKFKMKYWGVASFLQKGNDDHWIVD TDYDTYAVQYSCRL
LNL DGT CADSY SFVFSRDPNGLPPEAQKIVRQRQEELCLARQYRLIVHNGYCDGRSERNLL
```



**Choose Destination**

File  Clipboard  
 Collections  Analysis Tool

Download 1 items.

Format  
FASTA

Create File

Protein 3D Structure

# 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
MKVWVALLLLAALGSGRAERDCRVSSFRVKENFDKARFSGTWYAMAKKDP
EGLFLQDNIVAEFVDETGQMSATAKGRVRLNNDVDCADMVGTFTDTE
PAKFKMKYWGVASFQKGNDDHWIVDTDYDTYAVQYSCRLNLDGTCADS
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

<b>Official Symbol</b>	RBP4 <small>provided by HGNC</small>
<b>Official Full Name</b>	retinol binding protein 4, plasma <small>provided by HGNC</small>
<b>Primary source</b>	<a href="#">HGNC:9922</a>
<b>Locus tag</b>	PRO2222
<b>See related</b>	<a href="#">Ensembl:ENSG00000138207</a> ; <a href="#">HPRD:01580</a> ; <a href="#">MIM:180250</a> ; <a href="#">Vega:OTTHUMG00000018773</a>
<b>Gene type</b>	protein coding
<b>RefSeq status</b>	REVIEWED
<b>Organism</b>	<a href="#">Homo sapiens</a>
<b>Lineage</b>	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini; Catarrhini; Hominidae; Homo
<b>Also known as</b>	RDCCAS
<b>Summary</b>	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
  - RinAaccav

Note, links to many other RBP4 database entries are available

# Gene entry (links from genomic region)

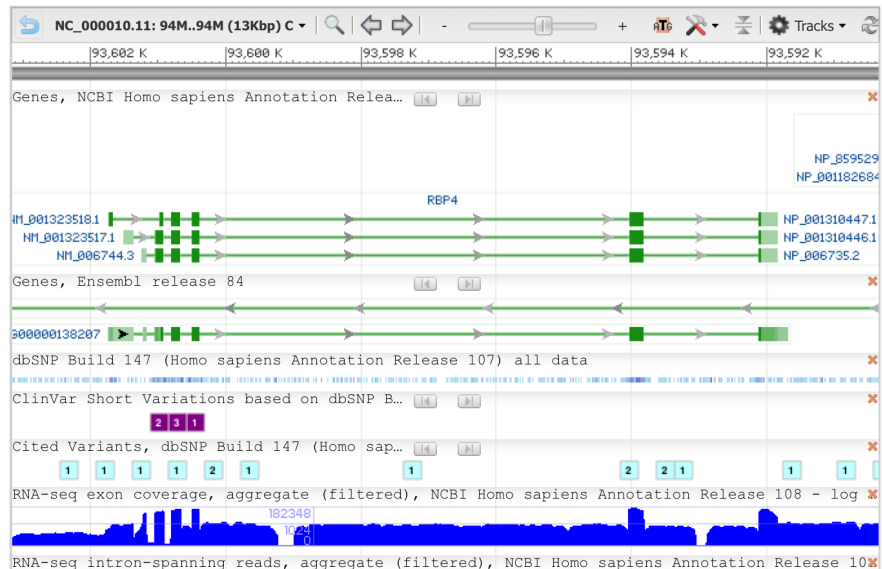


## Genomic regions, transcripts, and products

Go to [reference sequence details](#)

Genomic Sequence:  Chromosome 10 Reference GRCh38.p7 Primary Assembly

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



# Gene entry (links to related sequences)



## NCBI Reference Sequences (RefSeq)

RefSeqs maintained independently of Annotated Genomes

These reference sequences exist independently of genome builds. [Explain](#)

### Genomic

#### 1. [NG\\_009104.1](#) RefSeqGene

Range: 5001..14401  
Download: [GenBank](#), [FASTA](#), [Sequence Viewer \(Graphics\)](#)

### 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](#)  
UniProtKB/Swiss-Prot: [P02753](#)  
Related: [ENSP00000360522](#), [OTTHUMP0000020116](#), [ENST00000371467](#), [OTTHUMT00000049431](#)

#### 2. [NM\\_001323518.1](#) → [NP\\_001310447.1](#) retinol-binding protein 4 isoform b

Status: REVIEWED

Source sequence(s): [AL356214](#), [BC020633](#), [BG565176](#), [BI712834](#)  
UniProtKB/Swiss-Prot: [P02753](#)

# Gene entry (links in all categories)



## 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

# NCBI Online Books



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.

**NCBI Bookshelf** provides free online access to books and documents in life science and healthcare.

# PubMed to find literature information



PubMed.gov  
US National Library of Medicine  
National Institutes of Health

PubMed

Search

[Advanced](#)

[Help](#)



## PubMed

PubMed comprises more than 30 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.

**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 25 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



PubMed search interface showing results for "retinol binding protein fish".

Search results: 1 to 20 of 100

1. [Proteomic changes in the liver of Channa striatus in response to high temperature stress.](#)  
Mahanty A, Purohit GK, Banerjee S, Karunakaran D, Mahanty S, Mahanty BP. Electrophoresis. 2016 Jul;37(12):1704-17. doi: 10.1002/elps.201500393. Epub 2016 May 9. PMID: 27058960 [Similar articles](#)

2. [Nmnat1-Rbp7 Is a Conserved Fusion-Protein That Combines NAD+ Catalysis of Nmnat1 with Subcellular Localization of Rbp7.](#)  
Chen H, Babino D, Schoenbichler SA, Arkhipova V, Töchterle S, Martin F, Huck CW, von Lintig J, Meyer D. PLoS One. 2015 Nov 30;10(11):e0143825. doi: 10.1371/journal.pone.0143825. eCollection 2015. PMID: 26618989 **Free PMC Article** [Similar articles](#)

3. [Structure of zebrafish IRBP reveals fatty acid binding.](#)  
Ghosh D, Haswell KM, Sprada M, Gonzalez-Fernandez F. Exp Eye Res. 2015 Nov;140:149-58. doi: 10.1016/j.exer.2015.08.026. Epub 2015 Sep 4. PMID: 26344741 [Similar articles](#)

Search details: ("retinol-binding proteins"[MeSH Terms] OR ("retinol-binding"[All Fields] AND "proteins"[All Fields]) OR "retinol-binding proteins"[All Fields])

# PubMed search (refinement)



PubMed.gov US National Library of Medicine National Institutes of Health

PubMed retinol binding protein fish Search

Create RSS Create alert **Advanced** Help

Article types: Clinical Trial, Review, Customize ...

Text availability: Abstract, Free full text, Full text

PubMed Commons: Reader comments, Trending articles

Publication dates: 5 years, 10 years, Custom range...

Species: Humans, Other Animals

Clear all Show additional filters

Format: Summary Sort by: Most Recent Send to Filters: Manage Filters

**Search results**

Items: 1 to 20 of 100 << First < Prev Page 1 of 5 Next > Last >>

[Proteomic changes in the liver of \*Channa striatus\* in response to high temperature stress.](#)  
1. Mahanty A, Purohit GK, Banerjee S, Karunakaran D, Mohanty S, Mohanty BP. *Electrophoresis*. 2016 Jul;37(12):1704-17. doi: 10.1002/elps.201500393. Epub 2016 May 9. PMID: 27058960 [Similar articles](#)

[Nmnat1-Rbp7 Is a Conserved Fusion-Protein That Combines NAD+ Catalysis of Nmnat1 with Subcellular Localization of Rbp7.](#)  
2. Chen H, Babino D, Schoenbichler SA, Arkhipova V, Töchterle S, Martin F, Huck CW, von Lintig J, Meyer D. *PLoS One*. 2015 Nov 30;10(11):e0143825. doi: 10.1371/journal.pone.0143825. eCollection 2015. PMID: 26618989 **Free PMC Article** [Similar articles](#)

[Structure of zebrafish IRBP reveals fatty acid binding.](#)  
3. Ghosh D, Haswell KM, Sprada M, Gonzalez-Fernandez F. *Exp Eye Res*. 2015 Nov;140:149-58. doi: 10.1016/j.exer.2015.08.026. Epub 2015 Sep 4. PMID: 26344741 [Similar articles](#)

**Titles with your search terms**

Carotenoid and retinoid transport to fish oocytes and egg [Mol Aspects Med. 2003]  
Developmental expression, tissue dis [Comp Biochem Physiol B Biochem...]

See more...

**Find related**

Database: Select

Find items

**Search details**

("retinol-binding proteins"[MeSH Terms] OR ("retinol-binding"[All Fields] AND "proteins"[All Fields]) OR "retinol-binding proteins"[All Fields])

Search See more...

# PubMed entry



PubMed.gov US National Library of Medicine National Institutes of Health

PubMed Search

Advanced Help

Format: Abstract Send to

*Exp Eye Res*. 2015 Nov;140:149-58. doi: 10.1016/j.exer.2015.08.026. Epub 2015 Sep 4.

**Structure of zebrafish IRBP reveals fatty acid binding.**

Ghosh D<sup>1</sup>, Haswell KM<sup>2</sup>, Sprada M<sup>3</sup>, Gonzalez-Fernandez F<sup>4</sup>.

**Author information**

**Abstract**

Interphotoreceptor retinoid-binding protein (IRBP) has a remarkable role in targeting and protecting all-trans and 11-cis retinol, and 11-cis retinal during the rod and cone visual cycles. Little is known about how the correct retinoid is efficiently delivered and removed from the correct cell at the required time. It has been proposed that different fatty composition at that the outer-segments and retinal-pigmented epithelium have an important role in regulating the delivery and uptake of the visual cycle retinoids at the cell-interphotoreceptor-matrix interface. Although this suggests intriguing mechanisms for the role of local fatty acids in visual-cycle retinoid trafficking, nothing is known about the structural basis of IRBP-fatty acid interactions. Such regulation may be mediated through IRBP's unusual repeating homologous modules, each containing about 300 amino acids. We have been investigating structure-function relationships of Zebrafish IRBP (zIRBP), which has only two tandem modules (z1 and z2), as a model for the more complex four-module mammalian IRBP's. Here we report the first X-ray crystal structure of a teleost IRBP, and the only structure with a bound ligand. The X-ray structure of z1, determined at 1.90 Å resolution, reveals a two-domain organization of the module (domains A and B). A deep hydrophobic pocket with a single bound molecule of oleic acid was identified within the N-terminal domain A. In fluorescence titrations assays, oleic acid displaced all-trans retinol from zIRBP. Our study, which provides the first structure of an IRBP with bound ligand, supports a potential role for fatty acids in regulating retinoid binding.

Published by Elsevier Ltd.

**KEYWORDS:** Interphotoreceptor matrix; Interphotoreceptor retinoid-binding protein; Oleic acid; RBP3; Retina; Visual cycle; X-ray structure; Zebrafish

PMID: 26344741 PMCID: PMC4624502 [Available on 2016-11-01] DOI: 10.1016/j.exer.2015.08.026

**Full text links**

ELSEVIER FULL-TEXT ARTICLE

**Save items**

Add to Favorites

**Similar articles**

Module structure of interphotoreceptor retinoid-binding protein [BMC Biochem. 2007]  
Cone outer segment and Müller microvilli pericellular matrices [Exp Eye Res. 2013]  
Purification of the full-length Xenopus interphotoreceptor retinoid [Mol Vis. 2007]  
**Review** Interphotoreceptor retinoid-binding protein (IRBP) [Mol Neurobiol. 1993]  
**Review** Interphotoreceptor retinoid-binding protein [Prog Clin Biol Res. 1991]

See reviews... See all...

**Cited by 1 PubMed Central article**

Fold conservation and proteolysis in

## PubMed search strategies



Try the tutorials (“How To” and “help” at top)

Use boolean queries (capitalize AND, OR, NOT)

**lipocalin AND disease**

Utilize search fields (author [au], 1<sup>st</sup> 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 Entrez information and external resources

## 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.



**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.

## NCBI 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.

## NCBI 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.

# NCBI 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.

# Genomes via Ensembl



[BLAST/BLAT](#) | [VEP](#) | [Tools](#) | [BioMart](#) | [Downloads](#) | [Help & Docs](#) | [Blog](#)

## Tools

[All tools](#)

## BioMart >

Export custom datasets from Ensembl with this data-mining tool

## BLAST/BLAT >

Search our genomes for your DNA or protein sequence

## Variant Effect Predictor >

Analyse your own variants and predict the functional consequences of known and unknown variants

## Search

All species  for

e.g. [BRCA2](#) or [rat 5:62797383-63627669](#) or [rs699](#) or [coronary heart disease](#)

## All genomes

-- Select a species --

- [View full list of all Ensembl species](#)
- [Edit your favourites](#)

## Favourite genomes



**Human**  
GRCh38.p12

[Still using GRCh37?](#)



**Mouse**  
GRCm38.p6

## Genomes via Ensembl



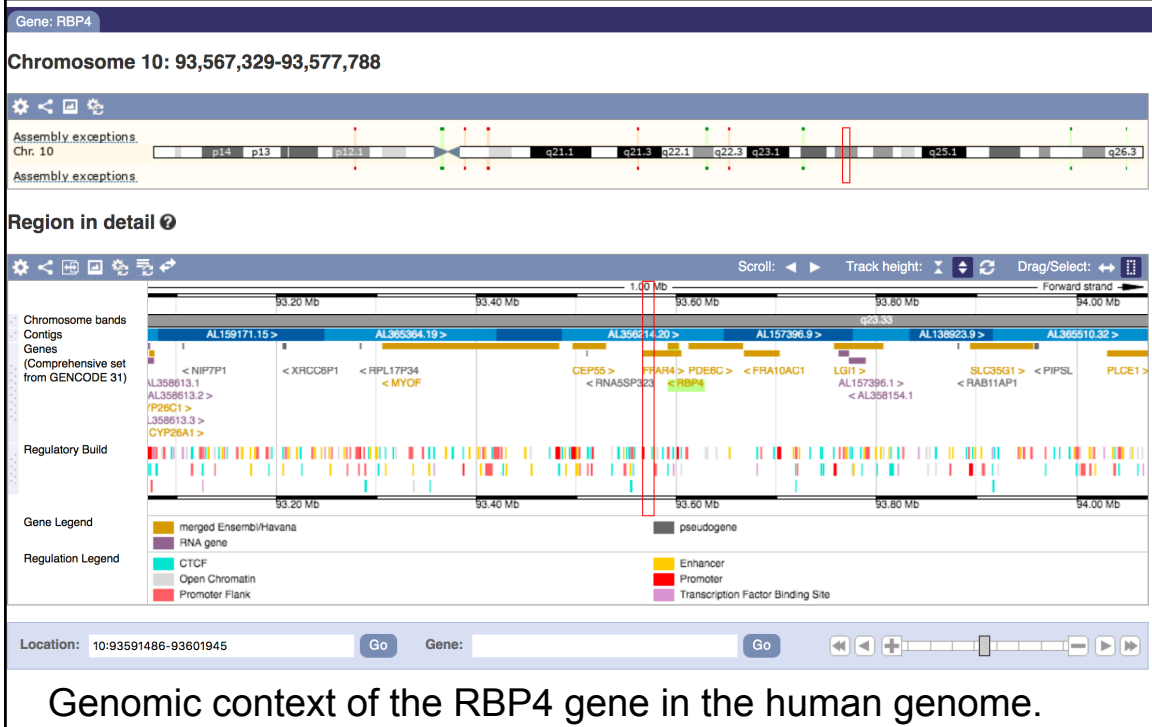
- Ensembl is a [genome annotation system](#), developed jointly by the EBI and the Wellcome Trust Sanger Institute.
- Originally used for the annotation, analysis and display of [vertebrate genomes](#) since 2000.
- Since 2009, complemented by the creation of five new sites, for [bacteria](#), [protists](#), [fungi](#), [plants](#) and [invertebrate metazoa](#).
- A single collection of interfaces for accessing and comparing genome-scale data from species of scientific interest from across the taxonomy.
- integrative power of Ensembl tools for [comparative analysis](#), [data mining](#) and [visualisation](#) across genomes.

## Tools at Ensembl/EnsemblGenomes



- [Searching Ensembl](#) – search for genes, keywords etc and have them displayed in the genomic context.
- [BLAST/BLAT](#) – sequence search against the genomes in Ensembl.
- [BioMart](#) – a data-mining tool to export custom datasets or conversion tables from Ensembl.
- [Variant Effect Predictor \(VEP\)](#) – determines the effect of your variants (SNPs, insertions, deletions, CNVs or structural variants) on genes, transcripts, and protein sequence, as well as regulatory regions.
- [Downloads](#) – Ensembl provides means to download the whole genomic data, i.e. genomic/transcript/peptide sequences (FASTA format), feature annotation files (GFF+GTF format), etc.

# Genome Viewing at Ensembl



# Genome Viewing – more details

