H3ABioNet

H3ABioNet

Pan African Bioinformatics Network for H3Africa

Introduction to Bioinformatics Online Course : IBT

Introduction to Databases and Resources

Biological Databases and Resources

Learning Objectives

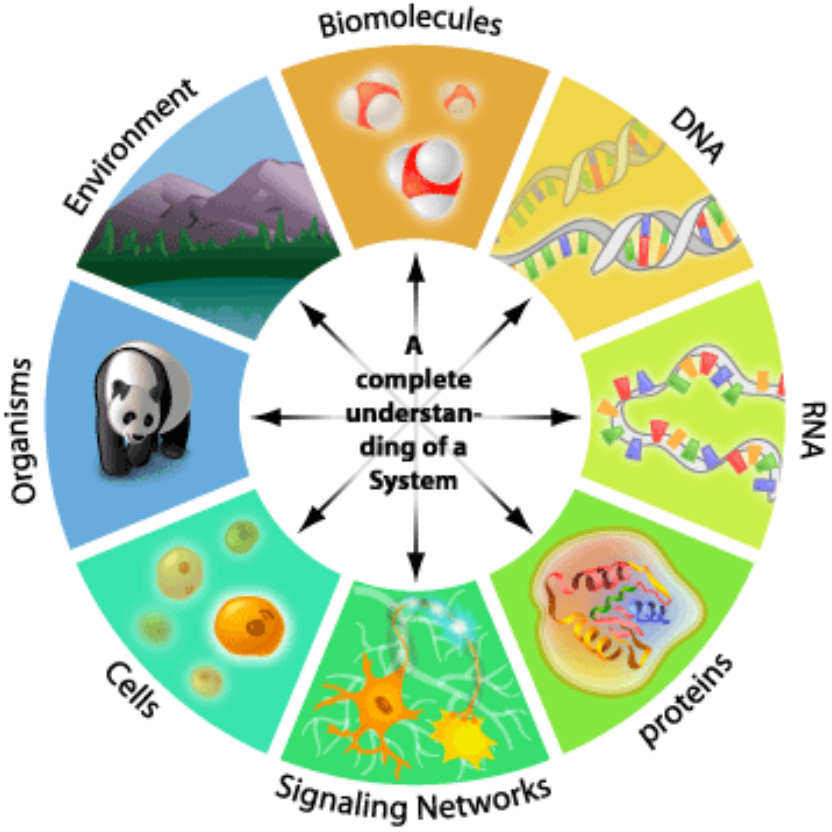
- Introduction to Databases and Resources
  - Understand how bioinformatics data is stored and organised
  - Describe the different types of data found at the NCBI and EBI resources
  - Locate key bioinformatics databases and resources

Learning Outcomes

- Introduction to Databases and Resources

  - Understand the structure and layout of the NCBI and EBI data resources
  - Understand the difference between databases, tools, repositories
  - Search for data from specific databases using accessions numbers, gene name
  - Use selected tools at NCBI and EBI

Data



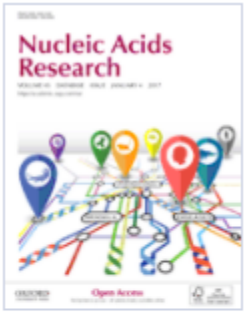
Introduction

- Range of different online databases and resources

- Need to know which:
  - Which databases and resources exist
  - What tools are available to mine these resources
  - What tools are available to search across resources

Biological databases

Nucleic Acids Research



The 24th annual *Nucleic Acids Research* database issue: a look back and upcoming changes FREE

Volume 45, Issue D1

January 2017

ISSN 0305-1048
EISSN 1362-4962

Volume 44, Issue W1

08 July 2016

ISSN 0305-1048
EISSN 1362-4962

Nucleic Acids Research: VOLUME 44 WEB SERVER ISSUE JULY 8, 2016 FREE

FRONT-MATTER/BACK-MATTER

Editorial

Web Server issue

Nucleic acid sequence, structure, and regulation

Protein sequence and structure, motifs, and domains

Metabolic and signalling pathways, enzymes

Viruses, bacteria, protozoa and fungi

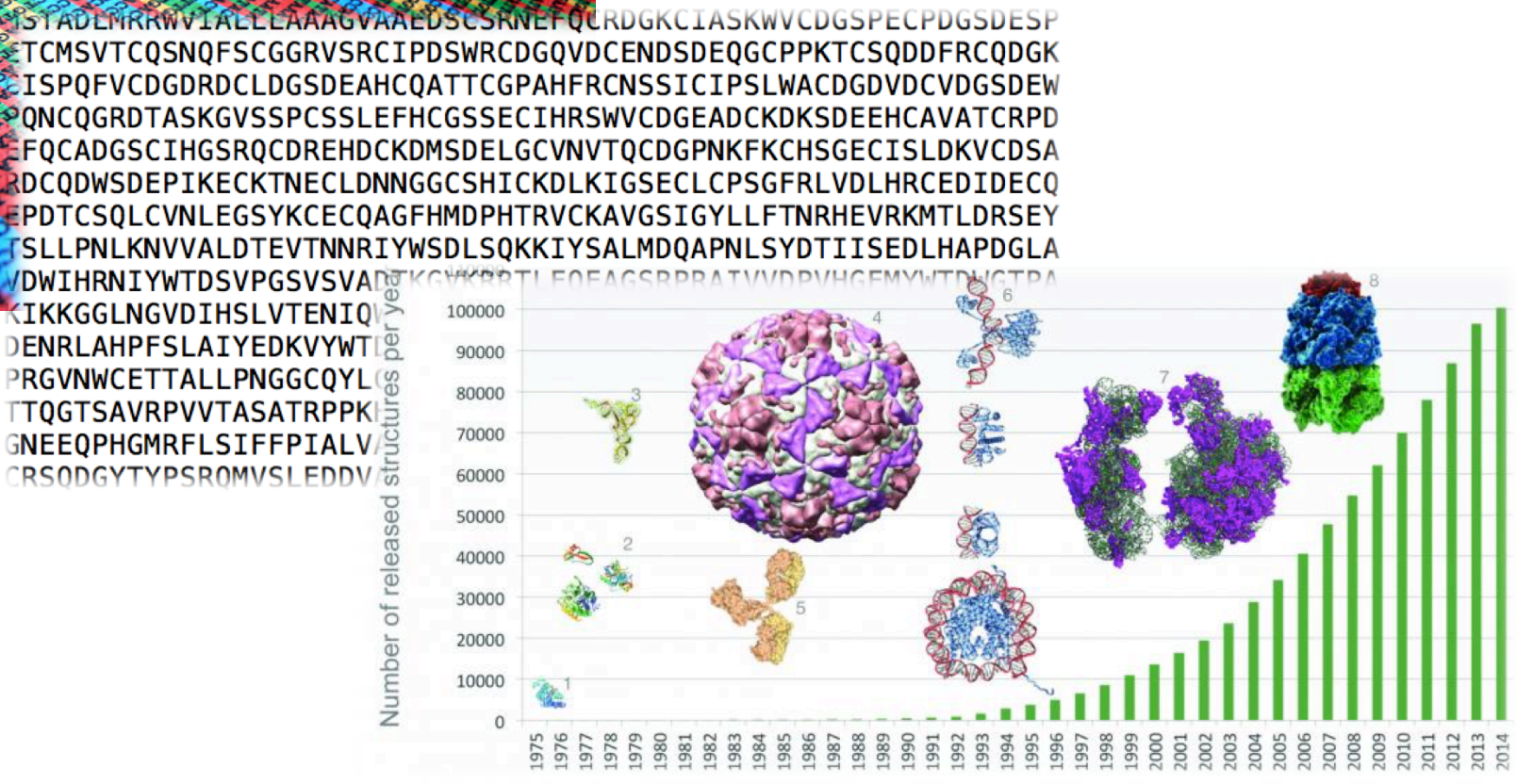
Human genome, model organisms, comparative genomics

Genomic variation, diseases, and drugs

Databases

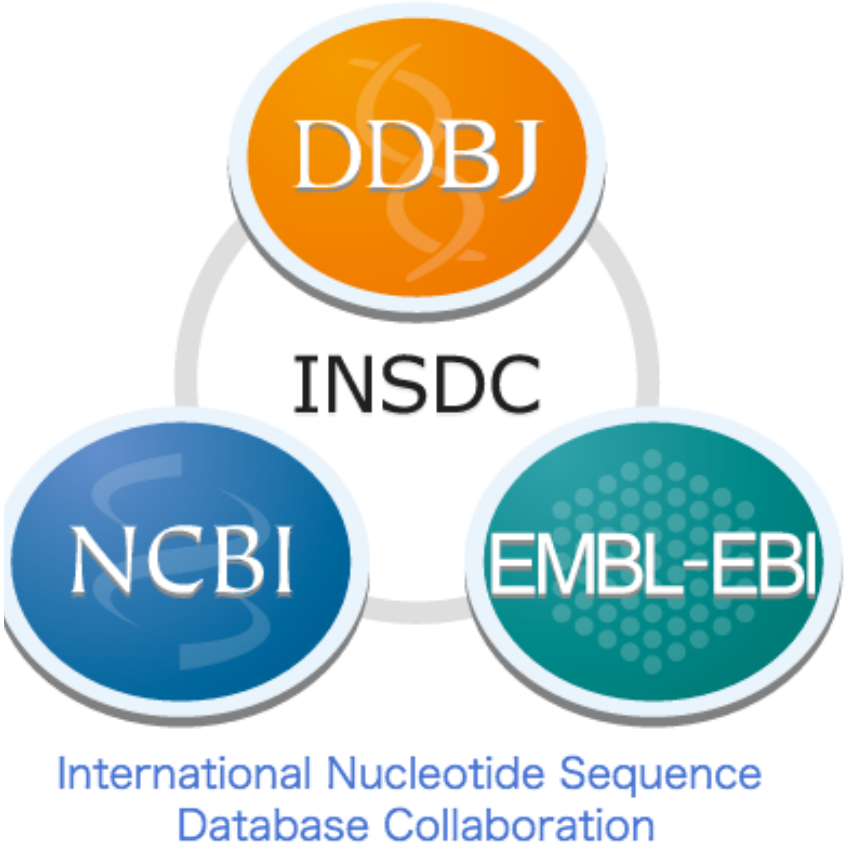
- Databases are:
  - Public or private
    - Access and submission
  - Protein, nucleotide, structure, literature, annotation…
  - Generalised or specialised
  - Curated or non-curated
  - Sequence or genome-centred

Primary Databases



Primary Databases

- International Nucleotide Sequence Database Collaboration (INSDC)
- Genomic sequence data stored in 3 public databases
- Each have own accession numbers and tools



Secondary Databases

- In-depth databases built upon primary sequence data
- Provide several different resources and annotations

Most Popular Bioinformatics Resources

- National Centre for Biotechnology Information (NCBI)

- European Bioinformatics Institute (EMBL-EBI)

NCBI

- National Centre for Biotechnology Information (NCBI)
  - National Institute of Health funded initiative established to store molecular biology information
  - Has grown dramatically since the completion of the human genome project
  - Developed and maintained a variety of databases and resources

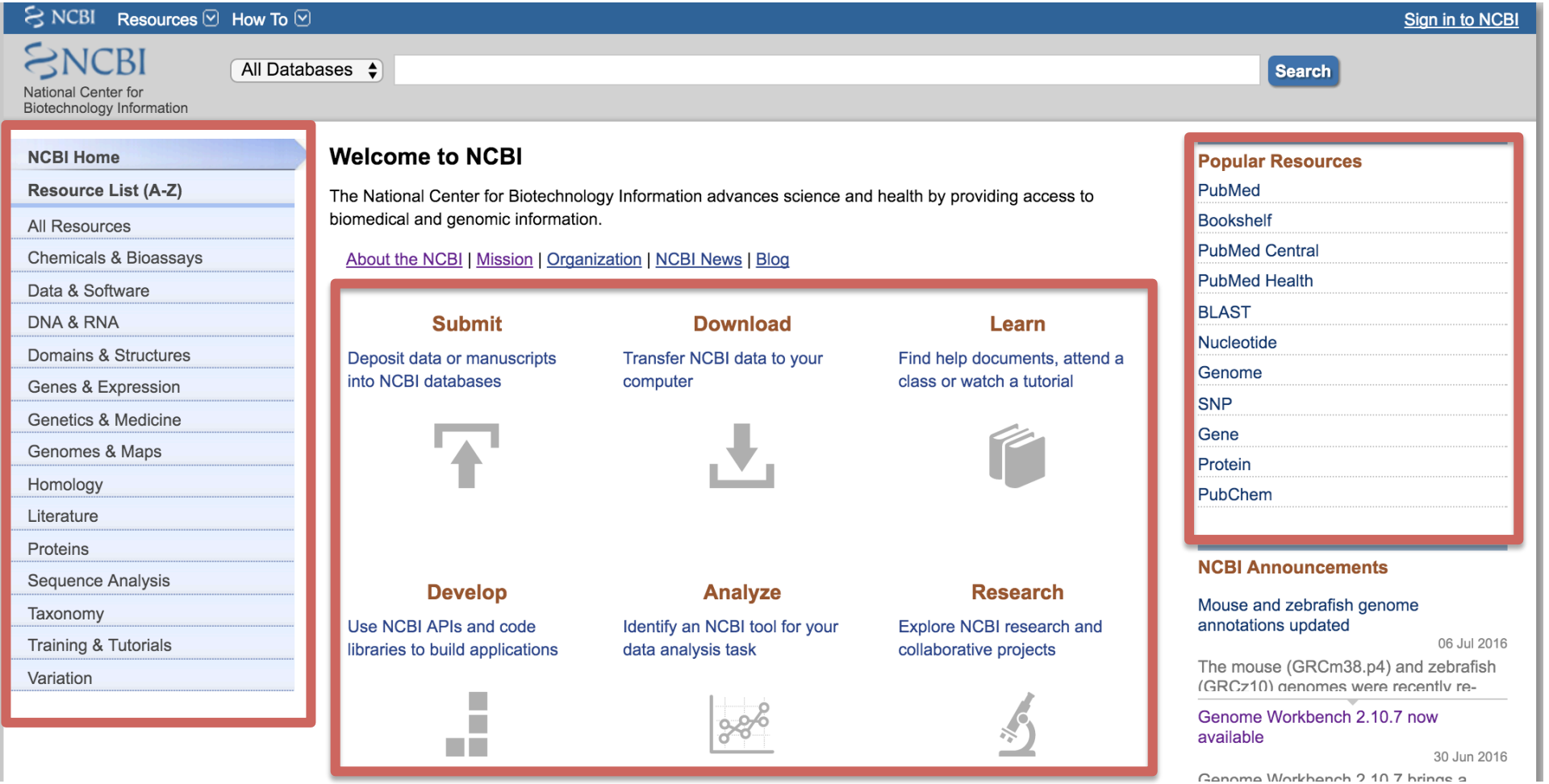
GenBank

- The NIH genetic sequence database
  - Contains an annotated collection of all publicly available DNA sequences
  - Part of INSDC
  - The database is updated on a regular basis, approximately every two months
  - Several divisions within GenBank

GenBank Divisions

- Expressed sequence tags (ESTs)
  - Short sub-sequences of a cDNA sequence
- High-throughout Genomic Sequences (HTGs)
  - Clone based HTGs
- Complete Microbial Genomes
- Whole Genome Shotgun Sequences(WGS)
- Transcriptome Shotgun Assembly Sequences (TSA)

NCBI



Analysis Tools

Analyze

NCBI provides a wide variety of data analysis tools that allow users to manipulate, align, visualize and evaluate biological data.

Selected Analysis Tools

All Tools | Literature | Health | Genomes | Genes | Proteins | Chemicals

Filter this table

Tools	Description
1000 Genomes Browser	Graphically depicts variant calls, genotype calls and supporting evidence (such as aligned sequence reads) that have been produced by the 1000 Genomes Project
Assembly Archive	Links the raw sequence information found in the Trace Archive with assembly information found in GenBank/EMBL/DDBJ
Basic Local Alignment Search Tool (BLAST)	Finds regions of local similarity between biological sequences
BLAST Microbial Genomes	Finds regions of local similarity between query sequences and sequences from complete microbial genomes
Electronic PCR (e-PCR)	Identifies sequence tagged sites (STSs) within DNA sequences
Gene Plot	Displays pairs of protein homologs that are symmetrical best hits between two genomes
Genome BLAST	Finds regions of local similarity between query sequences and genome sequences
Genome ProtMap	Maps each protein from a COG or VOG back to its genome

Tutorials

Learn

NCBI creates a variety of educational products including courses, workshops, webinars, training materials and documentation. NCBI educational events are free and open to everyone. All NCBI educational materials are available for anyone to re-use and distribute.

Webinars & Courses

In-person courses, live webinars and webinar recordings

Conferences & Presentations

Booth exhibits and workshops at scientific conferences

Tutorials

Tutorials: Training materials in HTML, PDF and video formats

Documentation

Online manuals, handbooks, fact sheets and FAQs

UPCOMING EVENTS

How to upload and analyze dbGaP data in the Cloud

FEBRUARY 3, 2016

Online Webinar: 1:00-2:00pm

Five ways to submit next-gen sequence data to NCBI's Sequence Read Archive

FEBRUARY 17, 2016

Online Webinar: 1:00-2:00pm

"NCBI Resources for Patent Searchers" at the PIUG Biotechnology 2016 Conference

FEBRUARY 24, 2016

Workshop

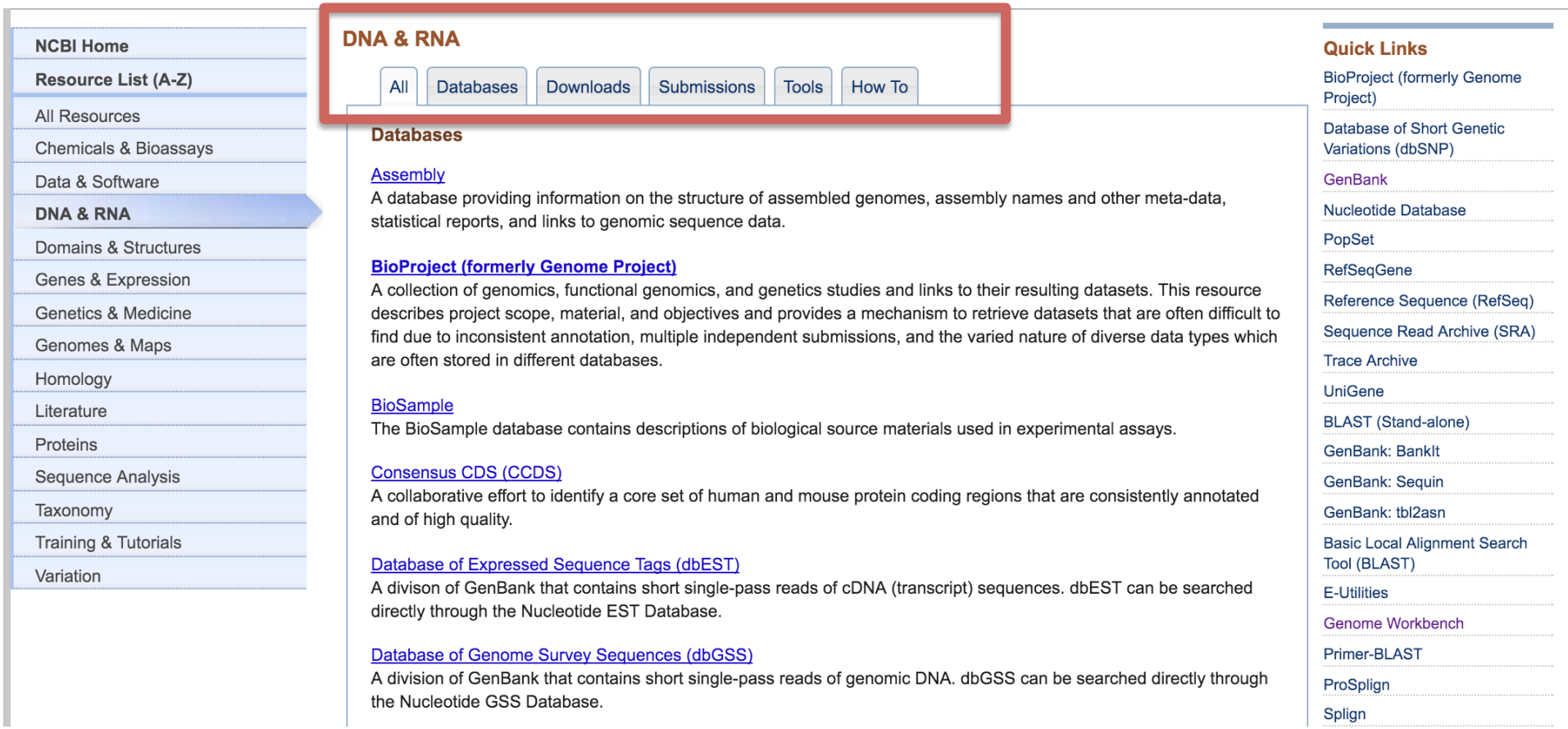
A Librarian's Guide to NCBI

MARCH 7-11, 2016

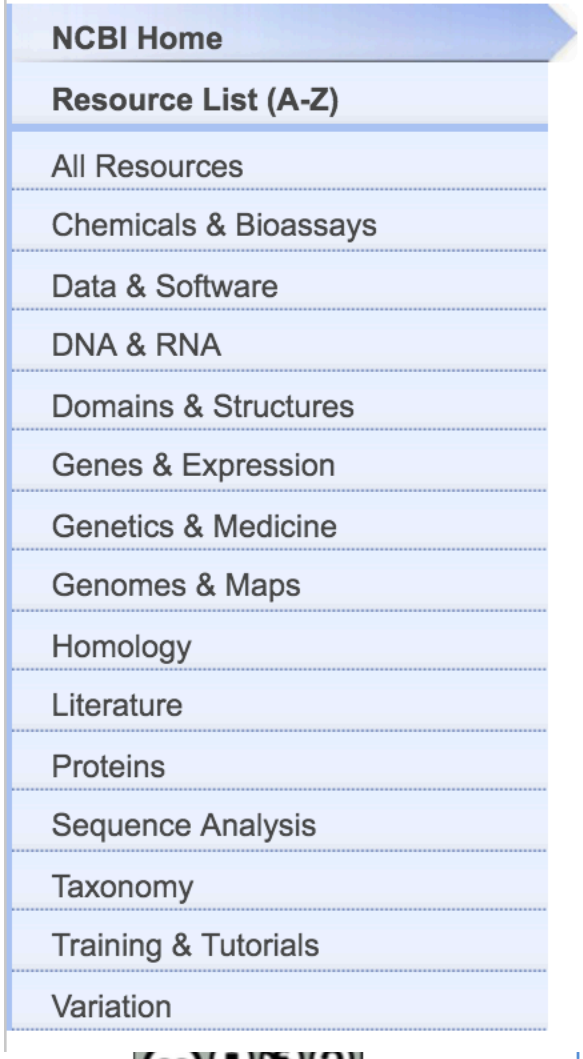
Workshop

Experimental Biology 2016 Annual Meeting

NCBI – DNA and RNA



NCBI – Not only DNA data



EMBL - EBI

- Maintain the world's most comprehensive range of freely available and up-to-date molecular databases
- Offer online and live training events for using their resources
  - https://www.ebi.ac.uk/training

EMBL – EBI

Bioinformatics services

We maintain the world's most comprehensive range of **freely available** and up-to-date molecular databases. Developed in collaboration with our colleagues worldwide, our services let you share data, perform complex queries and analyse the results in different ways. You can work locally by downloading our data and software, or use our web services to access our resources programmatically. You can read more about our services in the journal *Nucleic Acids Research*.

DNA & RNA
genes, genomes & variation

Gene expression
RNA, protein & metabolite expression

Proteins
sequences, families & motifs

Structures
Molecular & cellular structures

Systems
reactions, interactions & pathways

Chemical biology
chemogenomics & metabolomics

Ontologies
taxonomies & controlled vocabularies

Literature
Scientific publications & patents

Cross domain
cross-domain tools & resources

Popular

- Ensembl
- UniProt
- PDBe
- ArrayExpress
- ChEMBL
- BLAST
- Europe PMC
- Reactome
- Train online
- Support

Service news

EMBL – EBI

Services > DNA & RNA

Popular services

Ensembl

Ensembl enables and advances genome science by providing high-quality, integrated annotation on vertebrate genomes within a consistent and accessible infrastructure.

Ensembl Genomes

An integrating portal for genome-scale data from non-veterbrate species.

Clustal Omega

Multiple sequence alignment of DNA or protein sequences. Clustal Omega replaces the older ClustalW alignment tools.

European Nucleotide Archive

An open, supported platform for the management, sharing, integration, archiving and dissemination of public-domain sequence data.

Quick links

- Popular services in this category
- All services in this category
- Project websites in this category

ACCESSING DATA

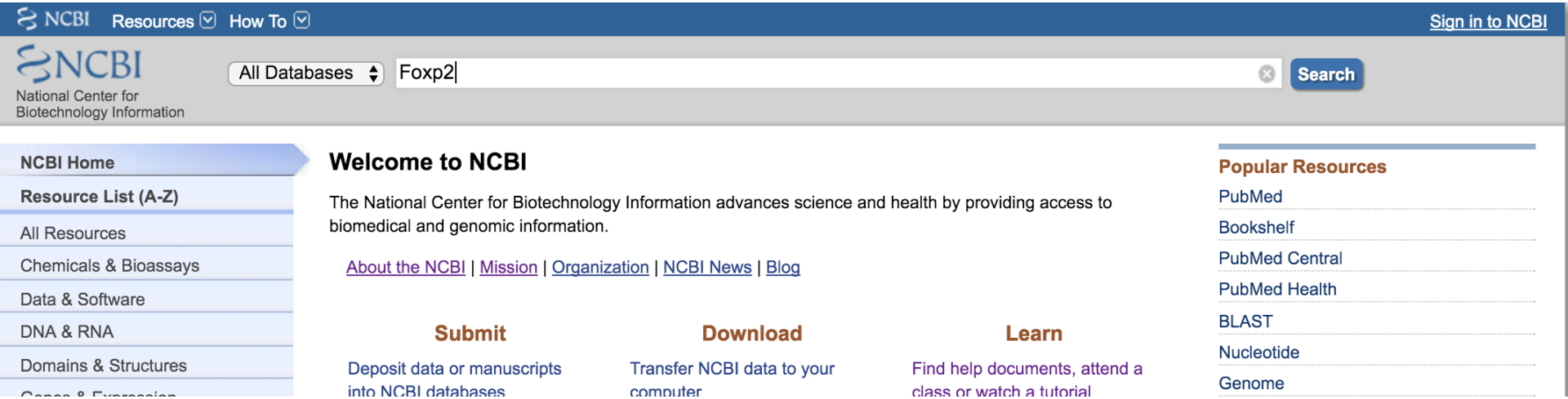
Accessing Data

- Why would you need to access sequence data?
  - Know what the sequence of a gene is
  - Identify variants in the sequence
  - Compare your sequence to others
  - Identify similar sequences
  - Find diseases associated with variation in your gene of interest

Accessing Data

- Important to be clear what data you are searching for
- Most tools have been developed to link to all annotations for a particular query
- Both NCBI and EBI provide portals to allow you to search across all available databases with a single query

Example: FOXP2 Human



NCBI Portal Search

Search NCBI databases Help

Literature

Genes

Health

Protein

Genomes

Chemicals

Results found in 31 databases for "Foxp2"

Literature

Books	68	books and reports
MeSH		
NLM Catalog		
PubMed	426	scientific & medical abstracts/citations
PubMed Central	1,653	full-text journal articles

Health

ClinVar	41	human variations of clinical significance
dbGaP	0	genotype/phenotype interaction studies
GTR		
MedGen		
OMIM	14	online mendelian inheritance in man
PubMed Health	0	clinical effectiveness, disease and drug reports

Genomes

Assembly	0	genome assembly information
BioProject	13	biological projects providing data to NCBI
BioSample	6	descriptions of biological source materials
Clone	2,715	genomic and cDNA clones
dbVar	644	genome structural variation studies
Genome	12	genome sequencing projects by organism
GSS	3	genome survey sequences
Nucleotide	1,672	DNA and RNA sequences
Probe	2,139	sequence-based probes and primers

Genes

EST	3	expressed sequence tag sequences
Gene	227	collected information about gene loci
GEO DataSets	124	functional genomics studies
GEO Profiles	9,430	gene expression and molecular abundance profiles
HomoloGene		
PopSet	11	studies
UniGene	13	clusters of expressed transcripts

Proteins

Conserved Do…		
Protein		
Protein Clusters	0	sequence similarity-based protein clusters
Structure	3	experimentally-determined biomolecular structures

Chemicals

BioSystems	235	molecular pathways with links to genes, proteins and chemicals
PubChem BioAssay	2	bioactivity screening studies
PubChem Compound	0	chemical information with structures, information and links
PubChem Substance	102	deposited substance and chemical information

Popular Databases

- **Gene** – One stop resource for all annotation information for a gene
- **PubMed** – Extensive biomedical literature database
- **Nucleotide** – Database of all DNA sequence data
- **SNP** – Database of single nucleotide polymorphisms
- **Protein** – Database of protein sequences

Popular Databases

- **RefSeq** – Comprehensive, integrated, well-annotated set of reference sequences – genomic, transcript and protein
- **OMIM** – Online Mendelian Inheritance in Man - Database of human genes and genetic phenotypes
- **ClinVar** – Database of genomic variation and the relationship to human health

Gene | Foxp2

Create alert | Advanced

Tabular | 20 per page | Sort by Relevance | Send to:

Did you mean Foxp2 as a gene symbol?
Search Gene for Foxp2 as a symbol.

Search results

Items: 1 to 20 of 225

<< First | < Prev | Page 1 of 12 | Next > | Last >>

See also 2 discontinued or replaced items.

clear

Name/Gene ID	Description	Location	Aliases	MIM
FOXP2 ID: 93986	forkhead box P2 [*Homo sapiens* (human)]	Chromosome 7, NC_000007.14 (114086310..114693772)	CAGH44, SPCH1, TNRC10	605317
Foxp2 ID: 114142	forkhead box P2 [*Mus musculus* (house mouse)]	Chromosome 6, NC_000072.6 (14901349..15441977)	2810043D05Rik, AI449000, CAG-16, D0Kist7	
foxp2 ID: 555242	forkhead box P2 [*Danio rerio* (zebrafish)]	Chromosome 4, NC_007115.6 (6414782..6630967, complement)		
Foxp2 ID: 500037	forkhead box P2 [*Rattus norvegicus* (Norway rat)]	Chromosome 4, NC_005103.4 (41364441..41942782)	RGD1559697	
FOXP2 ID: 751769	forkhead box P2 [*Taeniopygia guttata* (zebra finch)]	Chromosome 1A, NC_011463.1 (25370715..25773703, complement)		
FOXP2 ID: 449627	forkhead box P2 [*Pan troglodytes* (chimpanzee)]	Chromosome 7, NC_006474.4 (118440345..119044173)		
FOXP2 ID: 613237	forkhead box P2 [*Macaca mulatta* (Rhesus monkey)]	Chromosome 3, NC_027895.1 (139602847..140196871)		
foxp2	forkhead box P2 [*Xenopus*		cagh44, spch1, tnrc10	

Gene Database – Foxp2

Full Report

Send to:

FOXP2 forkhead box P2 [*Homo sapiens* (human)]

Gene ID: 93986, updated on 3-Jul-2016

- **Summary**
- **Genomic context**
- **Genomic regions, transcripts, and products**
- **Bibliography**
- **Phenotypes**
- **Variation**
- **HIV-1 interactions**
- **Interactions**
- **General gene information**
- **General protein information**
- **NCBI Reference Sequences (RefSeq)**
- **Related sequences**
- **Additional links**

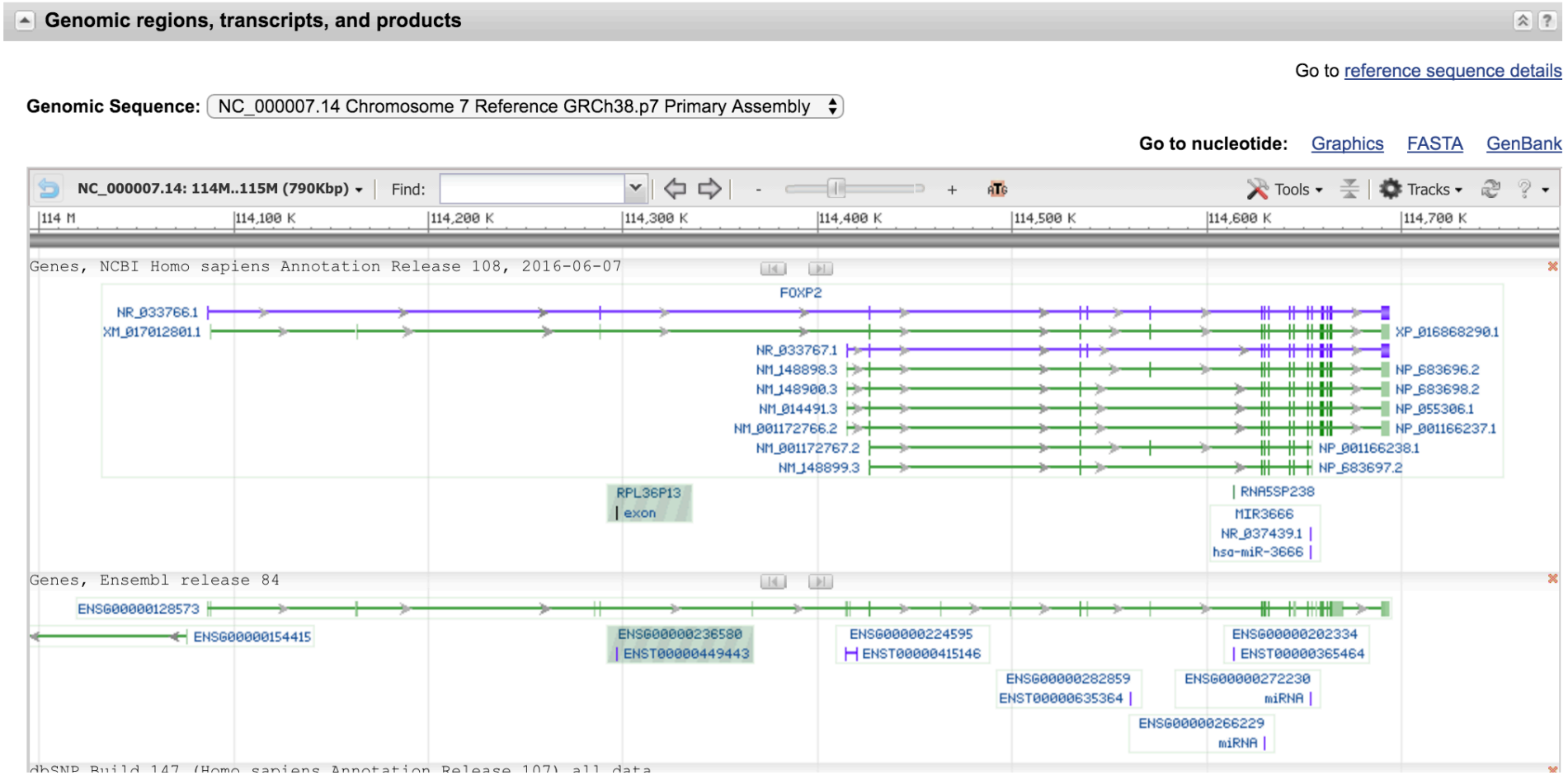
Gene Database – Foxp2

Gene ID: 93986, updated on 3-Jul-2016

Summary

Official Symbol	FOXP2 provided by HGNC
Official Full Name	forkhead box P2 provided by HGNC
Primary source	HGNC:HGNC:13875
See related	Ensembl:ENSG00000128573 HPRD:05611; MIM:605317; Vega:OTTHUMG00000023131
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	SPCH1; CAGH44; TNRC10
Summary	This gene encodes a member of the forkhead/winged-helix (FOX) family of transcription factors. It is expressed in fetal and adult brain as well as in several other organs such as the lung and gut. The protein product contains a FOX DNA-binding domain and a large polyglutamine tract and is an evolutionarily conserved transcription factor, which may bind directly to approximately 300 to 400 gene promoters in the human genome to regulate the expression of a variety of genes. This gene is required for proper development of speech and language regions of the brain during embryogenesis, and may be involved in a variety of biological pathways and cascades that may ultimately influence language development. Mutations in this gene cause speech-language disorder 1 (SPCH1), also known as autosomal dominant speech and language disorder with orofacial dyspraxia. Multiple alternative transcripts encoding different isoforms have been identified in this gene.[provided by RefSeq, Feb 2010]
Orthologs	mouse all

Gene Database – Foxp2



Gene Database – Foxp2

Bibliography

Related articles in PubMed

1. *FOXP2*-Related Speech and Language Disorders
   Morgan A, *et al.* , 1993. PMID 27336128
2. Downregulation of FOXP2 promoter human hepatocellular carcinoma cell invasion.
   Yan X, *et al.* Tumour Biol, 2015 Dec. PMID 26142732
3. Effect of pH on the Structure and DNA Binding of the FOXP2 Forkhead Domain.
   Blane A, *et al.* Biochemistry, 2015 Jun 30. PMID 26055196
4. Enhanced procedural learning of speech sound categories in a genetic variant of FOXP2.
   Chandrasekaran B, *et al.* J Neurosci, 2015 May 20. PMID 25995468, **Free PMC Article**
5. FOXF2 deficiency promotes epithelial-mesenchymal transition and metastasis of basal-like breast cancer.
   Wang QS, *et al.* Breast Cancer Res, 2015 Feb 26. PMID 25848863, **Free PMC Article**

See all (107) citations in PubMed

See citations in PubMed for homologs of this gene provided by HomoloGene

Gene Database – Foxp2

Phenotypes

Find tests for this gene in the NIH Genetic Testing Registry (GTR)

Review eQTL and phenotype association data in this region using PheGenI

Associated conditions

Description	Tests
Speech-language disorder 1 MedGen: C0750927, OMIM: 602081, GeneReviews: Not available	Compare labs

Copy number response

Description
Copy number response Triplosensitivity No evidence available (Last evaluated (2012-09-19) ClinGen Genome Curation Page Haploinsufficency Little evidence for dosage pathogenicity (Last evaluated (2012-09-19) ClinGen Genome Curation Page, PubMed

NHGRI GWAS Catalog

Description

Gene Database – Foxp2

Variation

See variants in ClinVar

See studies and variants in dbVar

See Variation Viewer (GRCh37.p13)

See Variation Viewer (GRCh38)

Genotypes

See SNP Geneview Report

See 1000 Genomes Browser (GRCh37.p13)

Gene Database– Foxp2

NCBI Reference Sequences (RefSeq)

RefSeqs maintained independently of Annotated Genomes

These reference sequences exist independently of genome builds. Explain

Genomic

1. **NG_007491.2 RefSeqGene**

Range	5001..612463
Download	GenBank, FASTA, Sequence Viewer (Graphics)

mRNA and Protein(s)

1. NM_001172766.2 → NP_001166237.1 **forkhead box protein P2 isoform V**

See identical proteins and their annotated locations for NP_001166237.1

Status: REVIEWED

Description	Transcript Variant: This variant (5) lacks an in-frame exon and uses an alternate in-frame splice site in the coding region, compared to variant 2. The resulting isoform (V) is shorter than isoform II.
Source sequence(s)	AC020606, AF337817, AI369947, BC018016, BC143867
UniProtKB/Swiss-Prot	O15409
UniProtKB/TrEMBL	B7ZLK5
Related	ENSP00000377135, OTTHUMP00000067771, ENST00000393498, OTTHUMT00000139941

Conserved Domains (1) summary

cd00059 Location:503 → 575	FH; Forkhead (FH), also known as a "winged helix". FH is named for the Drosophila fork head protein, a transcription factor which promotes terminal rather than segmental development. This family of transcription factor domains, which bind to B-DNA as ...

GenBank Entry– Foxp2

GenBank ▾ Send: ▾

Homo sapiens forkhead box P2 (FOXP2), RefSeqGene on chromosome 7

NCBI Reference Sequence: NG_007491.2

FASTA Graphics

Go to:

```
LOCUS       NG_007491              607463 bp    DNA     linear   PRI 04-JAN-2015
DEFINITION  Homo sapiens forkhead box P2 (FOXP2), RefSeqGene on chromosome 7.
ACCESSION   NG_007491 REGION: 5001..612463
VERSION     NG_007491.2  GI:299522979
KEYWORDS    RefSeq; RefSeqGene.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini;
            Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 607463)
  AUTHORS   Stroud JC, Wu Y, Bates DL, Han A, Nowick K, Paabo S, Tong H and
            Chen L.
  TITLE     Structure of the forkhead domain of FOXP2 bound to DNA
  JOURNAL   Structure 14 (1), 159-166 (2006)
   PUBMED   16407075
REFERENCE   2  (bases 1 to 607463)
  AUTHORS   Gauthier J, Joober R, Mottron L, Laurent S, Fuchs M, De Kimpe V and
            Rouleau GA.
```

GenBank Entry – Foxp2

Homo sapiens forkhead box P2 (FOXP2), RefSeqGene on chromosome 7

NCBI Reference Sequence: NG_007491.2

GenBank Graphics

```
>gi|299522979:5001-612463 Homo sapiens forkhead box P2 (FOXP2), RefSeqGene on
chromosome 7
AGACAGCGCGAGCCTCCGAGAAAGCGCGAGACACGCCGGCGCGTGCAGCTCCGCGGCCGCCGCTTCGCCC
TAGCTCTAGCCCCGCGCCACCCGCAGCCCGCCCGCGAACGCCCGCCCCGGTTATTTATGCGGCGGCCGCG
TCCGCTGGCTGCGGCTTCCTCGGCCCCCCCCCTCCCCGGGCGCGCCCCCCACTCGCGGCAGCAGCTGCCC
GGACTCGCGCGTGGGTGTGTTGTTTGGGGGCTTCTGCCTCGCCGCCGCGGGTGCCACCTCCCGGGACGCT
GCCCACGGCGTCCCCGGTCGCGGTAAGTTTCTTGGCCCTCACTCTGGCGCGCTACACCTCCGCACCCCAC
CCTGTCCCAGCCACCTCCACGCTGGGCCGAGCTGCGACTTTACTCTGCTCCGCGCTCCTCCCGCGGTGGC
GACAAAGTTTCGCCCCAAAGGCAGCGCCCTGCTTGCCGGGGCGAGTGTGACATGTGTCAAATTTGGGCTC
GGCGTTGGGGTCGCATTCCGGACCCAGCATCACCACTTGTTGTTCTCTTTCATTTTGCTTTGTGATGGGG
GAAAAAGGGGTGGAGAGGGGAGATTTGCTGTTGGCTCACGGATGATTTTTAGTTTGGATAATGCAAAATG
TTGCTTCGTCCGGAGAGACCTCGGCTGGAGGAGAATGTGTGTCGAAACACCAGATGTGTGTTTGTTTACT
CTCTTCTTTTTAATTGTTTGTTGTCTTTTCCCCTTCCTCCCCGCACCCCCACCCCCAACCTCGGGAGGAG
AAACAACAGTAAAAACATCTGGCGGTTAGAAGCACACACTTTATTGATCCAACTGCTGACCTTTATTTAC
TCAGTTGGGCAAGTGCACGCTTCTCGCGTCTAAGTTGGGCACTTCAGCGTTCATCTCAGAAGTACTTCTC
CAGGAAGGAGAGAGATGGAAAGGGGACACTCCTGTTTCTGGAGTCAAGAAACTCCTGGGCCCTACTGACG
CTTCGGGATAACCGTGCACAGGGATGACTGCTGCCTTGATCGCGTTCCTTTCCCTGTCCACGCGTTAGCA
CGACCGGCTTCCCCCGGTGTCTGGCCTGGTTTTACTTTTATTCCGCTTTAGAGAGGTCGCCTGGCTGTTT
GTGGGTGGGTTTGGGGTGAAATCGCACCTCGCGGCACTTGGTGAGGGGACGTGGGAGGAGCGCAGACACC
TTTCGGTGATAGGGAGGGCTTCTCACTTGGCTGTTACCTGGAAGTCCACAGTGGCCCCGGCGGGCAGGCG
GGCGGGCAGAGCGCGGGTCCGAACGCCCTTCGCGCTCGCGGCGGCGCACGTGCGGCGGCGGCGGCGGCGC
CGGCGGCGCGGGCGGGACCCACTGGGTGGCCGCGCGCGCCACCCGCGCGTTTTCTGCGCGCCTCTGCCCA
CCCGCGAGGCCGCAACGCGCCCTGACACCCGACCTCAGTGTGGACCCTCACTTGCTGTGGGTGTTGGGGC
CTCTCTAGAGCAGGGAGGAAAAGTTACCCTCACTATTCTCAGACTCTGGAATCTCTTAGGTAAGTCTCTT
AGGGCACCTGGCGCATGGGGTGACGCTTTGATTAGAAGGAACGTGAGGGGAAGAACGGGTGCAAGTCTGG
```

Accession Numbers

- Each GenBank record, consisting of both a sequence and its annotations is assigned a unique identifier called an accession number

GenBank ▾ Send: ▾

Homo sapiens forkhead box P2 (FOXP2), RefSeqGene on chromosome 7

NCBI Reference Sequence: NG_007491.2

FASTA Graphics

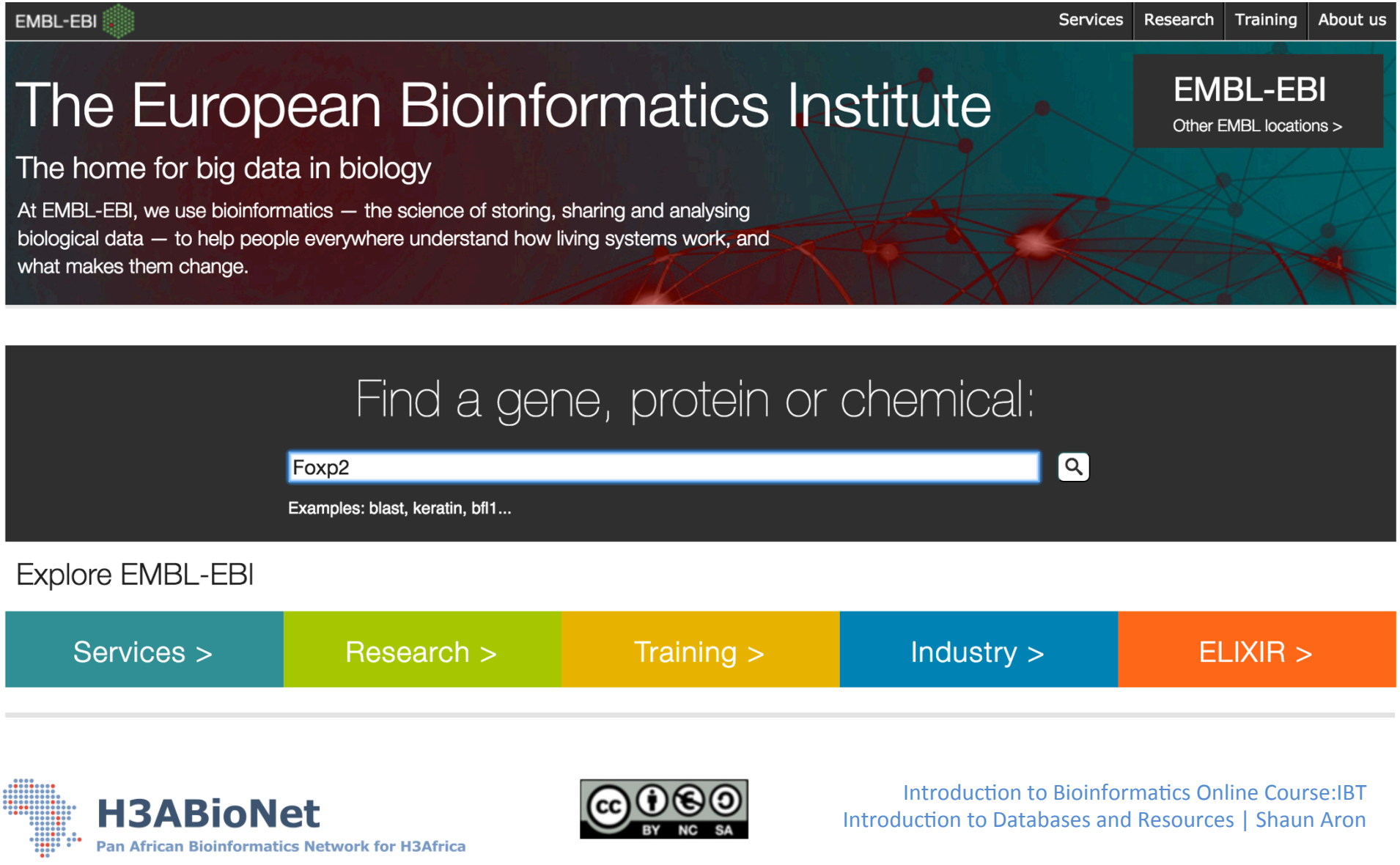
Go to:

```
LOCUS       NG_007491              607463 bp    DNA     linear   PRI 04-JAN-2015
DEFINITION  Homo sapiens forkhead box P2 (FOXP2), RefSeqGene on chromosome 7.
ACCESSION   ACCESSION   NG_007491 REGION: 5001..612463
VERSION     VERSION     NG_007491.2  GI:299522979
KEYWORDS
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini;
            Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 607463)
```

Accession Number Prefixes

Accession prefixes	Type	Description
NC_, NG_	Known Refseq	Genomic regions or assembly
NM_	Known Refseq	mRNA
NR_	Known Refseq	RNA
NP_	Known Refseq	Protein
NT_, NW_	Known Refseq	Genomic contig or scaffold
XM_	Model Refseq	mRNA
XR_	Model Refseq	RNA
XP_	Model Refseq	Protein

EMBL-EBI



EMBL – EBI – Foxp2

EBI Search

Help & Documentation | About EBI Search

Search results for *Foxp2*

Showing **17** results out of **2,885** in All results

Filter your results

Source

All results (2,885)
Genomes & metagenomes (71)
Nucleotide sequences (1,035)
Protein sequences (1,205)
Macromolecular structures (3)
Gene expression (28)
Molecular interactions (39)
Reactions, pathways & diseases (14)
Protein families (1)
Literature (388)
Samples & ontologies (100)
EBI web (1)

Gene & protein summaries (includes e

Forkhead box P2
FOXP2 (606354, TNRC10, SPCH1, CAGH44
Human (Homo sapiens)

Forkhead box P2
Foxp2 (CAG-16, D0Kist7, 2810043D05Rik,
House Mouse (Mus musculus)

View all entries in this group...
There are 2 more available.

View all available Gene & protein summaries

All results (2,885)
Genomes & metagenomes (71)
Nucleotide sequences (1,035)
Protein sequences (1,205)
Macromolecular structures (3)
Gene expression (28)
Molecular interactions (39)
Reactions, pathways & diseases (14)
Protein families (1)
Literature (388)
Samples & ontologies (100)
EBI web (1)

Nucleotide sequences (1,035 results found)

AFN11569
Eospalax baileyi **FOXP2**

Related data

Source: Coding (F
ID: AFN11569

EMBL – EBI – Foxp2

forkhead box P2

Gene Information and Sequence

- FOXP2 spans 607446 bps of chromosome 7 from 114086327 to 114693772.
- FOXP2 has 26 transcripts containing a total of 98 exons on the forward strand.
- Annotation for this gene includes both automatic annotation from Ensembl and Havana manual curation, see article.
- View the gene sequence in Ensembl.
- View the chromosome region for this gene in Ensembl.

Variations

- FOXP2 has 27589 SNPs.
- View sequence variations such as polymorphisms, along with genotypes and disease associations in Ensembl.

Orthologues

- FOXP2 has 66 orthologues in Ensembl
- View homology between species inferred from a gene tree in Ensembl.

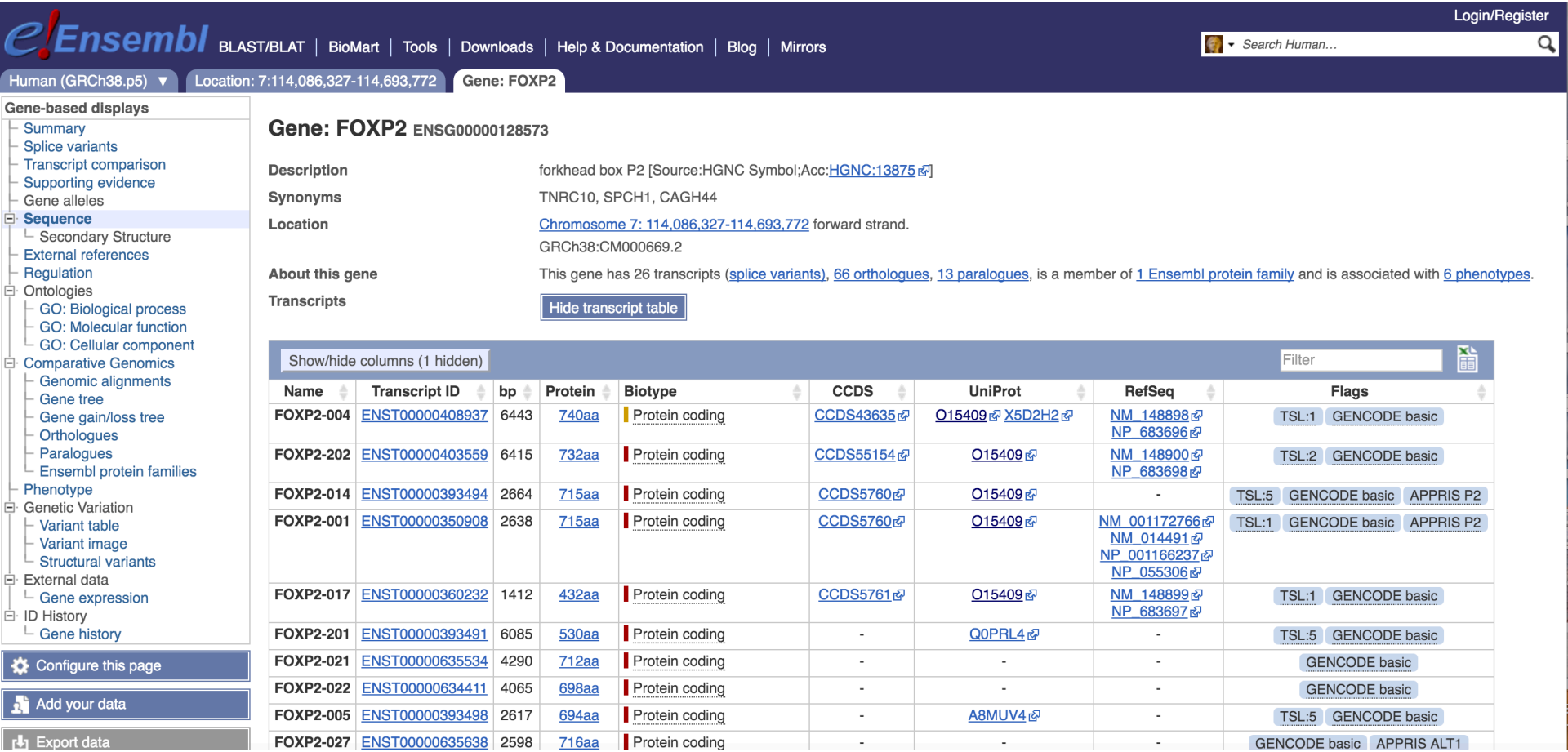
Paralogues

- FOXP2 has 13 paralogues in Ensembl
- View homology arising from a duplication event, inferred from a gene tree in Ensembl.

Regulation

- There are 47 regulatory elements located in the region of FOXP2.
- View the gene regulatory elements, such as promoters, transcription binding sites, and enhancers in Ensembl.

EMBL – EBI – Foxp2



EMBL – EBI – FoxP2

Gene
Expression
Protein
Protein Structure
Literature

FOXP2 expression summary

View in Expression Atlas

organism part

cultured skin substitute
View all

disease

chronic rhinosinusitis, poorly differentiated hypernephroma, pilocytic astrocytoma, ovarian cancer, ductal carcinoma in situ, normal, sonic hedgehog group medulloblastoma, Klinefelter's Syndrome, melanoma, intraductal papillary-mucinous adenoma (IPMA), adenocarcinoma of lung, adenocarcinoma of colon, Astrocytoma, Pilocytic, atypical teratoid/rhabdoid tumor, adenocarcinoma prostate PC-3 subline isolated from liver metastasis in mice, atypical teratoid / rhabdoid tumor, intraductal papillary-mucinous neoplasm (IPMN), invasive ductal carcinoma, glioblastoma, intraductal papillary-mucinous carcinoma (IPMC)
View all

Homo sapiens

Gene
Expression
Protein
Protein Structure
Literature

Forkhead box protein P2

View in UniProt

Subcellular Location

Nucleus.
View annotation in UniProt

Disease

Speech-language disorder 1 (SPCH1) [MIM:602081]: A disorder characterized by severe orofacial dyspraxia resulting in largely incomprehensible speech. Affected individuals have severe impairment in the selection and sequencing of fine orofacial movements which are necessary for articulation, and deficits in several facets of grammatical skills and language processing, such as the ability to break up words into their constituent phonemes. Note=The disease is caused by mutations affecting the gene represented in this entry.

EMBL – EBI – FoxP2

Gene
Expression
Protein
Protein Structure
Literature

2 protein structures available

Structure of the DNA binding domains of NFAT and FOXP2 bound specifically to DNA.

View in PDBe

Method

X-ray diffraction

Experiment

Resolution: 2.7Å
R-Factor: 23.82%
Free R-Factor: 28.72%

Dates

Deposited: 22-08-2005
Released: 08-08-2006
Revised: 24-02-2009

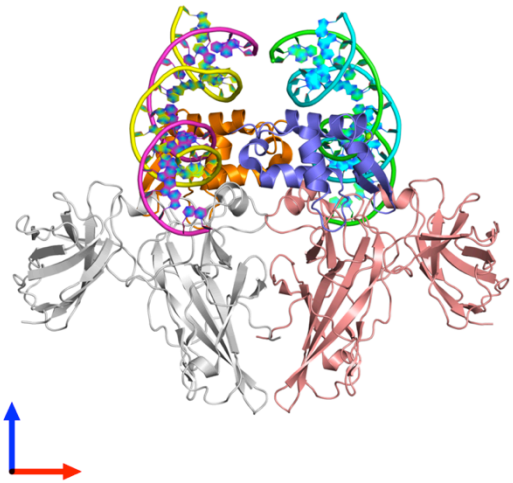
Deposited by

Wu, Y.,Stroud, J.C.,Borde, M.,Bates, D.L.,Guo, L.,Han, A.,Rao, A.,Chen, L.

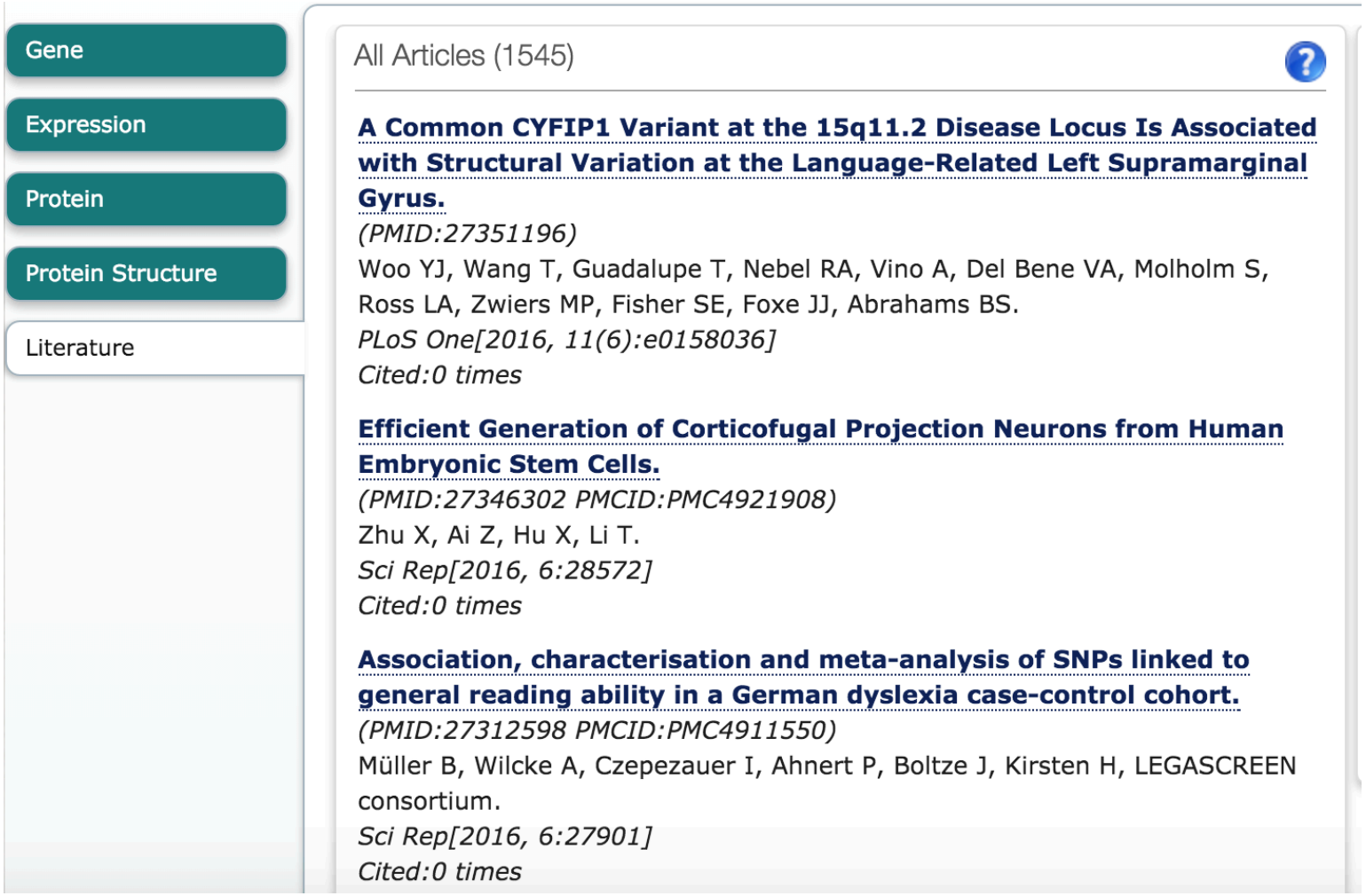
Primary Citation

FOXP3 controls regulatory T cell function through cooperation with NFAT.
Cell vol:126 page:375-87 (2006)
View citation in PDBe

Macromolecules



EMBL – EBI – FoxP2



Other popular resources at EBI

- **Ensembl** – resource for high quality intergrated annotation data
- **Uniprot** – Universal Protein Resource for protein sequence and functional annotation data
- **PDBe** – Protein data bank Europe – Collection of 3D structural data
- **InterPro** – database of protein families, domains and conserved sites

Specialised Databases

- A large number of specialised databases exist
  - Most of the sequences are also in GenBank/EMBL bank
  - May contain whole genomes
  - May contain specialised resources
  - Contain specific tools for mining the data

Specialised Databases

- Plasmodium http://plasmodb.org
- Sanger's specialised collections http://www.sanger.ac.uk
- Hepatitis Database http://hcv.lanl.gov/content/hcv-db/index
- Influenza Research Database http://www.fludb.org/

Summary

- Large amount of data out there
- Primary databases store raw sequence data
- Secondary databases provide information on the annotation of the sequence data
- Important to know how and where data is stored
- NCBI and EBI are the two most popular resources for extracting biological data