



March 30 – April 1, 2015

Alfred D. Hershey Building, Cold Spring Harbor Laboratory

Schedule

Day 1 (Monday)

7:00 pm – 9:30 pm: Course Introduction, Genome Sequencing and Assembly

- Course Introduction (King) (7:00-7:30)
- Genome Sequencing and Assembly (Dr. Michael Schatz, CSHL) (7:30-9:30)
- Social Hour (Blackford Hall) (9:30-10:30)

Day 2 (Tuesday)

9:00 – 12:30: Basic Browser and Database Resources

- Sequence, Gene, & Protein Resources (Ward) (9:00-10:00)
- Break (10:00-10:15)
- UCSC Genome Browser (Gordon) (10:15-11:30)
- Ensembl (Howell) (11:30-12:30)

12:30 - 1:30 - Lunch

1:30 – 6:00: Advanced and Comparative Analyses

- Comparative Genome Analysis (Ward) (1:30-2:30)
- Functional Genomic Elements and ENCODE (King) (2:30-3:15)
- Break (3:15-3:30)
- Sequence Polymorphisms (Howell) (3:30-4:15)
- Overview of Next Generation Sequence Data Analysis (King) (4:15-5:15)
- Workshop (5:15-6:00)

6:00 – 8:00: Cocktail Reception and Dinner
in Clarkson Dining Room (Blackford Hall)

8:00 – 9:00: Research Seminar

Transposon Expression in Somatic Tissues
(Dr. Molly Hammell, CSHL)

Day 3 (Wednesday)

9:00 - 12:15: Focus on Next Generation DNA sequencing

- RNA-Seq Analysis (King) (9:00-10:00)
- Break (10:00-10:15)
- Sequence Analysis Using Galaxy (Gordon) (10:15-12:15)

12:15 - 1:15 - Lunch

1:15 - 5:00: Focus on RNAseq and Pathways

- RNA-Seq Analysis Using R (King) (1:15-2:30)
- Gene Set Enrichment and Pathway Analysis (Howell) (2:30-3:30)
- Break (3:30-3:45)
- Workshop (3:45-4:45)
- Course Summary (4:45-5:00)

The Genome Access Course

Introduction



The Genome Access Course

Cold Spring Harbor Laboratory

Ben King	Mount Desert Island Biological Laboratory
Gareth Howell	Jackson Laboratory
Assaf Gordon	New York Genome Center
Charla Lambert	Cold Spring Harbor Laboratory
Jeremy Ward	Middlebury College



April 2015

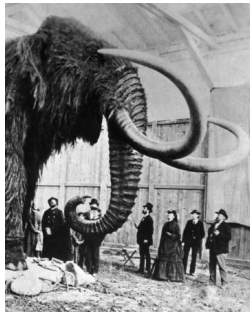
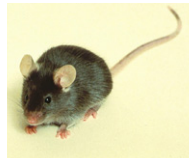
The \$1000 Genome Is Revolutionizing How We Study Biology and Practice Medicine

\$2.7B Over Decade → \$1000 in ~3 Days

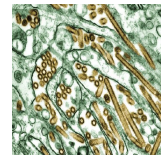


April 2015

Genomes, Genomes and More Genomes



Yellow fever and dengue fever mosquito vector



H5N1 Flu strains from around the world

April 2015

Human Genome Sequencing

- Reference human genome assembly
- Individual genomes
 - J. Craig Venter
 - Jim Watson
 - ... many others
- Your genome for \$9,500
 - Illumina's Personal Genome Service (<http://www.everygenome.com>)



1000 Genomes
A Deep Catalog of Human Genetic Variation

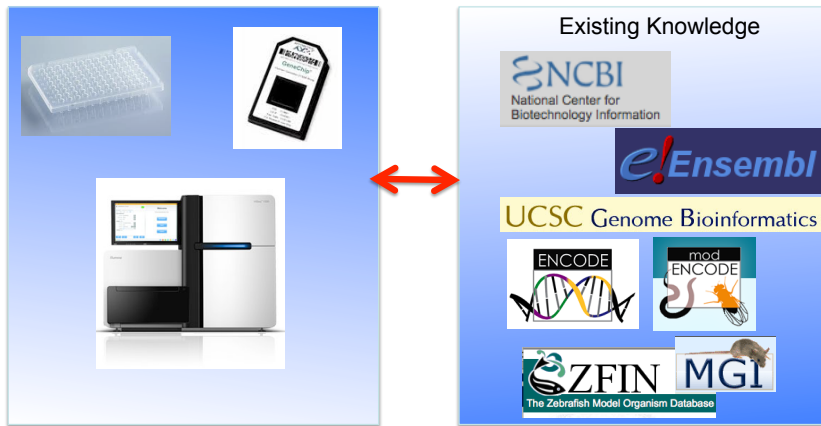


Genomics
england
100,000 Genomes Project

April 2015

Biology Increasingly Data Intensive

Need to leverage existing knowledge to advance your own research



April 2015

Analysis Techniques Becoming More Transparent

RESEARCH ARTICLE

CANCER

Detecting Early-Stage Tumors

Chetan Bette
Yuxuan Wang
Rhoda M. Ala
Henry Brem,
Peter Gibbs,
Matthias Hol
Giulia Sirave
Sue Kazue
Louise Olsson
le-Ming Shi
Jeanne Tie,
Christopher
Peter J. Allen,
Kenneth W. K

```
##### Table of Summary Statistics #####
data_table = data.frame(data)
cols_for_table <- c(rep(c(9), each=7), 10, 2, 8)
table <- generate.table(data_table, whichcols=cols_for_table, statistic=c(rep(c(1,2,5,5,5,5,5),1),rep(3,3)))
rows <- c("","Age","Mean (SD)","Median (Range)","No. Missing (%)",
,"Gender, N(%)", "Female","Male","Missing ",
,"Tumor Type, N(%)", "Bladder","Breast","Colorectal","Endometrial","Gastroesophageal","Glioma","Head and Neck Squamous Cell","Hepatoc",
,"Clinical Stage", "1","2","3","4","Primary Brain Tumors")
rownames(table) <- rows
colnames(table) <- c("All Patients")
write.csv(table,"sum_table_ST1_all.csv",na = "")
xt <- xtable(table, label="sum_table",caption="Table of Summary Statistics for All Patients")
align(xt) <- rep("t")
#digits(xt)[c(6)] <- 3
#print(xt, include.rownames=TRUE, scalebox=7, hline.after=c(-1,1,nrow(table)))
print(xt, include.rownames=TRUE, size="footnotesize", tabular.environment="longtable", floating=FALSE, hline.after=c(-1,1,nrow(table)))
```

Table S7 (Microsoft Excel format). Association between clinical characteristics and ctDNA concentration (log scale) in metastatic CRC patients.

Table S8 (Microsoft Excel format). Patient characteristics and plasma mutations detected after EGFR blockade.

Filename: generate.table.r
Filename: Table_S1_analysis.r
Filename: Table_S3_analysis.r

[Tables S1 to S8]

[r scripts]

Return to article

Science

19 February 2014

16 inspiring profiles of women making a difference in biology.

FREE Download

Created in partnership by

Science

FOR WOMEN IN SCIENCE

FOR WOMEN IN SCIENCE

FOR WOMEN IN SCIENCE

April 2015

Analysis of Genomic Data Not Only Critical, But Fun

Research

Open Access

Serendipitous discovery of *Wolbachia* genomes in multiple *Drosophila* species

Steven L Salzberg*, Julie C Dunning Hotopp*, Arthur L Delcher*,
Mihai Pop*, Douglas R Smith†, Michael B Eisen‡ and William C Nelson*

Address: *The Institute for Genomic Research, 9712 Medical Center Drive, Rockville, MD 20850, USA; †Agencourt Bioscience Corporation, 100 Cummings Center, Beverly, MA 01915, USA; ‡Center for Integrative Genomics, University of California, Berkeley, CA 94720, USA.

Correspondence: Steven L Salzberg, E-mail: salzberg@tifg.org

Published: 22 February 2005

Genome Biology 2005, 6:R23

The electronic version of this article is the complete one and can be found online at <http://genomebiology.com/2005/6/2/R23>

Received: 22 December 2004

Revised: 24 January 2005

Accepted: 24 January 2005



Wolbachia – endosymbionts found in fly embryos

Genome sequences for three new *Wolbachia*
species reconstructed from sequence traces from
7 *Drosophila* species

95% (1,440,650 base pairs) for one of the new species

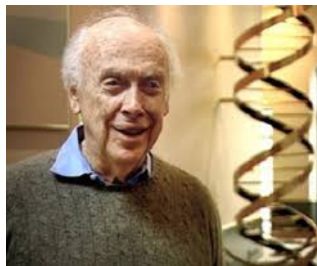
April 2015



The Genome
Access Course



Open Door
Workshops



James Watson



David Bentley

April 2015

The Genome Access Course

The Open Door Workshops

The Open Door Workshop provides an introduction to bioinformatics tools freely available on the internet, focusing primarily on the Human Genome data. The workshops provide hands-on training in the use of public databases and web-based sequence analysis tools, and are taught by experienced instructors.

The workshops are aimed at research scientists with a minimum of a degree in a biological discipline, including laboratory and clinical staff as well as specialists in related fields. Acceptance will be subject to selection process. The workshops run over three days and are without exception residential on the Wellcome Trust Genome Campus.

The Open Door Workshops were conceived jointly by The Wellcome Trust Sanger Institute and the Cold Spring Harbor Laboratory, and are being developed further by staff at these centres with assistance from staff at the European Bioinformatics Institute and the University of Cambridge.

The workshop is split into modules, each of which gives an overview of a given topic followed by a worked example and participant tasks.

The course is subsidised by the Wellcome Trust for scientists based in academic institutions anywhere in the world.

Workshops

- Kilifi, Kenya (3rd - 7th November 2014) - **CANCELLED**
- Kenya/Wellcome Trust Programme, Kilifi, Kenya
- Ref: ODW-Kenya
- Blantyre, Malawi (26th - 30th January 2013) - **CLOSED**
- Malawi-Liverpool-Wellcome Trust Clinical Research Programme, Blantyre, Malawi
- Ref: ODW-Malawi

Previous workshops

- Madison, United States (24th June 2014) - **Ended**
- The Madison Concourse Hotel
- Montevideo, Uruguay (26th - 27th March 2014) - **Ended**
- The Instituto de Higiene, Montevideo
- Workings with Encode Data Resources 1 and 2 (28th and 29th April 2014) - **Ended**
- The International Conference Centre Geneva
- Johannesburg, South Africa (10th - 13th October 2013) - **Ended**
- The Mallow Hotel, Sandton, Johannesburg
- Barcelona, Spain (5th July 2013) - **Ended**
- Polau de Congressos de Catalunya
- Singapore (Chander 16th April 2013) - **Ended**
- The Sands Expo and Convention Center
- Ha Chi Minh City, Vietnam (3rd - 5th December 2012) - **Ended**
- Oxford University Clinical Research Unit
- Edinburgh, Scotland (18th July 2011) - **Ended**
- Edinburgh International Conference Centre

Other open doors

- Uniquely manual
- Thailand manual

Related courses

- Other Wellcome Trust Advanced Courses
- The Genome Access Course at Cold Spring Harbor, Labs

April 2015

The Genome Access Course

The Genome Access Course

Aims:

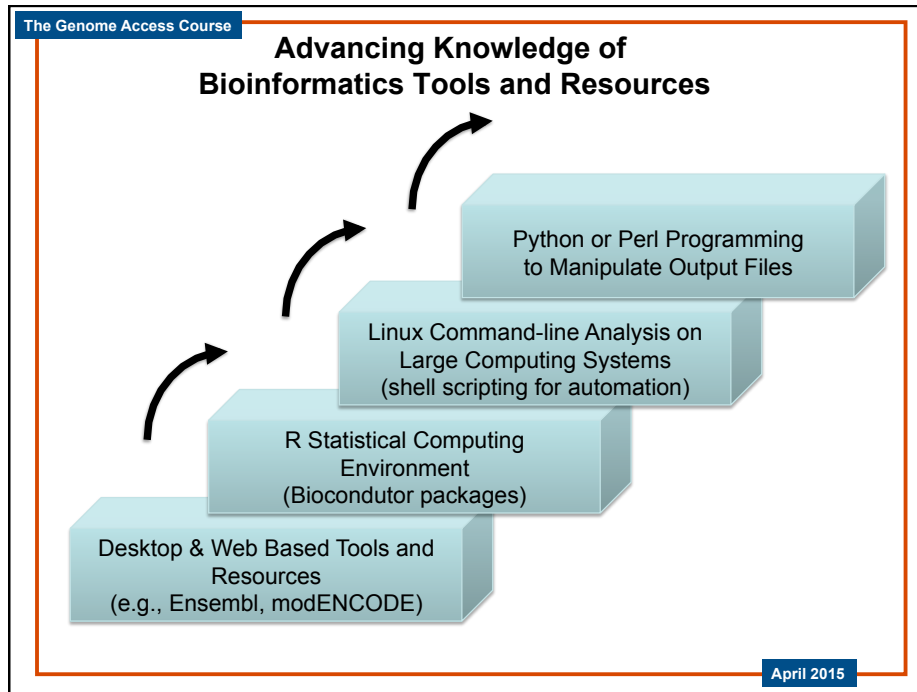
- What information is available
- How to access it
- What analysis tools are available

➔ Leave with knowledge of how to begin to apply these tools to your own research.



Statistical Methods for Functional Genomics (June 18-July 1, 2014)
 Computational Cell Biology (July 22-Aug. 11, 2014)
 Programming for Biology (Oct. 13-28, 2014)
 Advanced Sequencing Technologies & Applications (Nov. 11-23, 2014)
 Computational and Comparative Genomics (Nov. 5-18, 2014)

April 2015



The Genome Access Course		
Schedule Day 1 (Monday) 7:00 pm - 9:30 pm: Course Introduction, Genome Sequencing and Assembly - Course Introduction (King) (7:00-7:30) - Genome Sequencing and Assembly (Dr. Michael Schatz, CSHL) (7:30-9:30) - Social Hour (Blackford Hall) (9:30-10:30)		
Day 2 (Tuesday) 9:00 - 12:30: Basic Browser and Database Resources - Sequence, Gene, & Protein Resources (Ward) (9:00-10:00) - Break (10:00-10:15) - UCSC Genome Browser (Gordon) (10:15-11:30) - Ensembl (Howell) (11:30-12:30)	Day 3 (Wednesday) 9:00 - 12:15: Focus on Next Generation DNA sequencing - RNA-Seq Analysis (King) (9:00-10:00) - Break (10:00-10:15) - Sequence Analysis Using Galaxy (Gordon) (10:15-12:15)	12:15 - 1:15 - Lunch 1:15 - 5:00: Focus on RNAseq and Pathways - RNA-Seq Analysis Using R (King) (1:15-2:30) - Gene Set Enrichment and Pathway Analysis (Howell) (2:30-3:30) - Break (3:30-3:45) - Workshop (3:45-4:45) - Course Summary (4:45-5:00)
12:30 - 1:30 - Lunch 1:30 - 6:00: Advanced and Comparative Analyses - Comparative Genome Analysis (Ward) (1:30-2:30) - Functional Genomic Elements and ENCODE (King) (2:30-3:15) - Break (3:15-3:30) - Sequence Polymorphisms (Howell) (3:30-4:15) - Overview of Next Generation Sequence Data Analysis (King) (4:15-5:15) - Workshop (5:15-6:00)		
6:00 - 8:00: Cocktail Reception and Dinner in Clarkson Dining Room (Blackford Hall) 8:00 - 9:00: Research Seminar Transposon Expression in Somatic Tissues (Dr. Molly Hammell, CSHL)		

April 2015

Course Website

<https://courses.cshl.edu/>

Available after course – useful URLs and lecture slides

Course evaluations

April 2015

HIGH-THROUGHPUT BIOLOGY: FROM SEQUENCE TO NETWORKS

April 27 - May 3, 2015

Application Deadline: March 1, 2015

**In collaboration with
the Canadian Bioinformatics Workshops and
the New York Genome Center**

**This course will take place in New York City at the
main campus of the New York Genome Center
(NYGC).**

Lead Faculty:

Guillaume Bourque, Matei David, Marc Fiume, Francis
Ouellette, Jared Simpson, Malachi Griffith, Obi Griffith, Juri
Reimand, Quaid Morris, Irina Kalatskaya, Robin Haw, and
Michael Hoffman

http://meetings.cshl.edu/courses/2015/c-cbw15_nygc.shtml

April 2015

**STATISTICAL METHODS FOR
FUNCTIONAL GENOMICS**

June 18 - July 1, 2015

Application Deadline: March 15, 2015

Instructors:

Harmen Bussemaker, *Columbia University*
Sean Davis, *National Institutes for Health*
Olivier Elemento, *Weill Cornell Medical College*
Rafael Irizarry, *Dana-Farber Cancer Institute*

<http://meetings.cshl.edu/courses/2015/c-data15.shtml>

Programming environment: R, Bioconductor

Topics: Experimental design, quality control, normalization, statistical inference, correlation, multivariate statistics, multiple comparisons, false discovery rates, overrepresentation, reproducibility

April 2015

PROGRAMMING FOR BIOLOGY

October 12 - 27, 2015

Application Deadline: July 15, 2015

Instructor:

Simon Prochnik, *DOE Joint Genome Institute*
Sofia Robb, *Stowers Institute for Medical Research*

<http://meetings.cshl.edu/courses/2015/c-info15.shtml>

Programming environment: Perl, Bioperl

Topics: Introductory coding, biological libraries, analysis pipelines, data acquisition, validation, integration, analysis, and visualization

April 2015

**COMPUTATIONAL & COMPARATIVE
GENOMICS**

October 28 - November 3, 2015

Application Deadline: July 15, 2015

Instructors:

William Pearson, *University of Virginia, Charlottesville*

Lisa Stubbs, *University of Illinois, Urbana*

<http://meetings.cshl.edu/courses/2015/c-ecg15.shtml>

Programming environment: None

Topics: Sequence alignment, regulatory element and motif identification,
analysis of next-generation sequencing data, metagenomics

April 2015

**ADVANCED SEQUENCING
TECHNOLOGIES & APPLICATIONS**

November 10 - 22, 2015

Application Deadline: July 15, 2015

Instructors:

Elaine Mardis, *Washington University School of Medicine*

Gabor Marth, *University of Utah*

W. Richard McCombie, *Cold Spring Harbor Laboratory*

Aaron Quinlan, *University of Utah*

Michael Schatz, *Cold Spring Harbor Laboratory*

Co-Instructors:

Obi Griffith, *Washington University School of Medicine*

Malachi Griffith, *Washington University School of Medicine*

<http://meetings.cshl.edu/courses/2015/c-seqtec15.shtml>

Programming environment: UNIX

Topics: SNP calls, structural variant detection, read alignments, comparative
assembly, *de novo* assembly, cloud computing, Galaxy

April 2015

The Genome Access Course

Sequence, Gene and Protein Resources



ACCTTCACTTGAGCTAGGAAGGGCGCATTACCTTACTGGAGGTAAAGTAAGTTTGGCCTCTGTTG
 TCATGTGAGTCGTATCTAGAAACGCAGCAAGGGTGTGGGAATATCTTAACACAGGGTCTCATTGG
 AGCCCGTAGTTTACAGAAACCTCAGGCGGTCGGCATTACGGCTGTGACATTTGAGGTGAAGGTGT
 CAGCATATATCCACCCCCACCCCAACCCCTGGCTACTAAAGTAACATTGGTTTGTGTTTGTGAG
 ACAGGATCTTGGTGTGTAGTTGAGGCTGGCCTCAAAGCCTAGGGTGGCCTGGGACAAAGTCAGCC
 TTCTGAGTGTGGGATTAGAGGTGCGGGATATCAGTCACAGTGCCAGGAAACCCTTAGTACTAA
 GCACATGTGTATGCCGAGTCTCCTTTTCCTTGGTATATGATTGATACTCCCACTATCAAATACTG
 TAI ATCA
 CAG AAT
 TAC TCT
 ATA CTC
 CTG AAG
 TGT...GAAA
 CCCTGACTAAGACAGAGCTGTGCTTTTATAGTGAACACTTGAAAAATAACTCAGTCTAGAGAA
 TTTGTTGGAATGTTTACAATCAAACCTGTAGCAACATTGATAGATGTATTGTTTAACTGAATATA
 TAAACAGGCATTTTAAATGTAGAATTTTCAAGCTTCATAATAGATAAATTAGCCTCATCTAATTT
 TATTAGGCAGCTATTCTTTTAAAAATTAAGGCAGTTTGGGTAAGGGATATGAGACGTCAAGATGTA
 TAAAATAACCCAAAGATTATCACTTAAGATTTTACTGAAAAGATGCTGCAGGTATGTGCAGTCT
 CATGTTTGCCAAGTATGAGGTGCACTAGTCCCACTGTCTGAAATTCAGTGTAGAATGGCCTGA
 CCATGAGCTTTAAGAAGGTCTGTGTGTTTCTGCCTCCCAAGTACTGTGCAAGACTTGCATCAGCA
 TTCCTGGTCTTAGTAGAATGCTGTGTGTTTGGTGTGGAACATGGCCCCAAAGGCAGATTTAA
 TTTCTGTTTCTTGTACTATGTGATGTTTCTAAACATT

Sequence, Gene, and Protein Resources

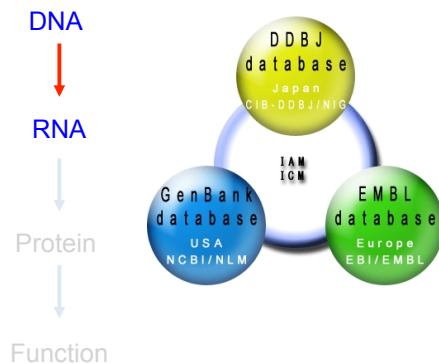
April 2015

Overview

- Accessing gene/protein information
 - NCBI Entrez system
 - UniProt
 - Gene Ontology
- RNA-Seq data repositories
 - NCBI Gene Expression Omnibus and EBI ArrayExpress
 - NCBI Short Read Archive and EBI European Nucleotide Archive

April 2015

There are three major nucleotide databases: DDBJ (Japan), EMBL (Europe), and GenBank (USA)



Any species (known or unknown)
Any type

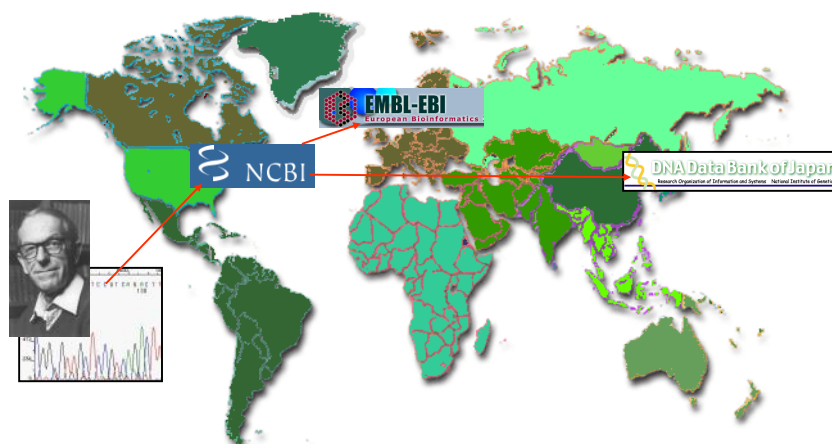
- **Genomic** - entire genes, entire chromosomes, fragments, BACs, constructs
- **cDNA** - full-length, partial, Expressed Sequence Tags (ESTs)
- **Other** - tRNA, ...

Any quality

- Draft vs. finished genomic sequences (e.g., BACs)
- ESTs vs. high quality cDNAs (e.g., mammalian gene collection)
- Highly Redundant

April 2015

Information is mirrored daily between DDBJ, GenBank and EMBL.



April 2015

RNA-seq data repositories



April 2015

Easiest Initial Way to Obtain Information

- **NCBI Entrez**
 - GenBank
 - PubMed
 - RefSeq
 - UniProt
 - Other Resources
- **Genome browser (e.g., Ensembl or UCSC)**
 - Query by chromosomal coordinates or gene name

Extensive links to other resources and also pre-computed BLAST searches

Then head to **secondary and organism specific databases**

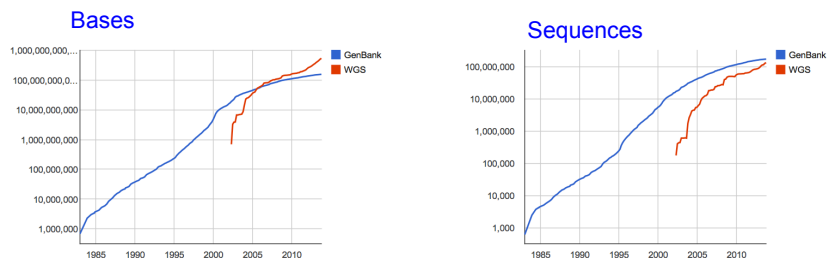
April 2015

ACCTTCACTTGAGCTAGGAAGGGCGCATTACCTTACTGGAGGTAAAGTAAGTTTGGCCTCTGTTG
 TCATGTGAGTCGTATCTAGAAACGCAGCAAGGGTGTGGGAATATCTTAACACAGGGTCTCATTGG
 AGCCCGTAGTTACAGAAACCTCAGGCGGTCGGCATTACGGCTGTGACATTTGAGGTGAAGGTGT
 CAGCATATATCCACCCCCACCCCAACCCCTGGCTACTAAAGTAACATTGGTTTGTGTTTTGAG
 ACAGGATCTTGGTGTGTAGTTGAGGCTGGCCTCAAAGCCTAGGGTGGCCTGGGACAAAGTCAGCC
 TTCTGAGTGTGGGATTAGAGGTGCGGGATATCAGTCACAGTGCCAAGGAAACCCTTAGTACTAA
 GCACATGTGTATGCCGAGTCTCCTTTTCCTTGGTATATGATTGATACTCCCACTATCAAATACTG
 TATGACCTGTTGGAGCTTGTACTGGCTGGTTTGTGTGTCAACTTGACACAGGCTGGGGTTATCA
 CAGAGAAAGGAGCT' GGTATTTTCTCAAT
 TACTGATCAAGGGG' TAGTCTTGGGTCT
 ATAAGAGAGGAGGC' CTCTGCATCAGCTC
 CTGCTTCTGACCT' AGCAATGTGAAAAG
 TGTAAAGCCGAATAA' TGCAGGAATAGAAA
 CCCTGACTAAGACAGAGCTGTTGCTTTTATAGTGAACACTTGAAAAATAA'CTCAGTCTAGAGAA
 TTTGTTGGAATGTTTACAATCAAAGTGTAGCAACATTGATAGATGTATTGTTTAACTGAATATA
 TAAACAGGCATTTTAAATGTAGAATTTTCAAGCTTCATAATAGATAAATTAGCCTCATCTAATTT
 TATTAGGCAGCTATTCTTTTAAATTAAGGCAGTTTGGGTAAGGGATATGAGACGTCAAGATGTA
 TAAAATAACCCAAAGATTATCACTTAAGATTTTACTGAAAAGATGCTGCAGGTATGTGCAGTCT
 CATGTTTGCCAAGTATGAGGTGCACTAGTCCCAACTGTCTGAAATTCACGTAGAATGGCCTGA
 CCATGAGCTTTAAGAAGGTCTGTGTGTTTCTGCCTCCCAAGTACTGTGCAAGACTTGCATCAGCA
 TTCTTGGTCTTAGTAGAATGCTGTGTGTTTGGTGTGGAACATGGCCCCAAAGGCAGATTTAA
 TTTCTGTTTCTTGTACTATGTGATGTTTCTAAACATT

The Basics: Gene Information

7
April 2015

GenBank, a historical repository, has doubled in size roughly every 18 months since 1982



- GenBank version 200 (Feb., 2014)
 - 171 million sequences (139 million WGS)
 - 158 billion NTs (591 billion NTs WGS)

8
April 2015

GenBank contains author and bulk submissions organized into 11 taxonomic divisions and 8 functional divisions

Author submitted

Bulk submitted

1. EST

2. GSS

3. WGS

4. HTG

US Patent Office

Table 1.
Growth of GenBank divisions (nucleotide base pairs)

Division	Description	Release 191 (8/2012)	Annual incn (%) ^a
Taxonomic divisions			
SYN	Synthetic	928 200 038	494.2%
PHG	Phages	84 079 451	34.4%
ENV	Environmental samples	3 374 433 548	32.1%
VRL	Viruses	1 429 464 786	21.1%
BCT	Bacteria	8 439 854 434	21.0%
PLN	Plants	5 481 470 133	15.6%
MAM	Other mammals	863 036 872	6.9%
VRT	Other vertebrates	2 886 594 595	6.7%
PRI	Primates	6 317 656 773	3.3%
UNA	Unannotated	127 803	1.5%
ROD	Rodents	4 435 106 948	0.9%
INV	Invertebrates	2 493 058 927	-1.7%
Functional divisions			
TSA	Transcriptome shotgun data	5 759 588 580	207.3%
WGS	Whole-genome shotgun data	308 196 411 905	47.9%
PAT	Patented sequences	12 118 622 726	8.6%
GSS	Genome survey sequences	21 947 780 105	5.7%
EST	Expressed sequence tags	40 888 051 100	4.8%
HTG	High-throughput genomic	24 359 210 558	0.1%
STS	Sequence tagged sites	636 262 446	0.1%
HTC	High-throughput cDNA	639 165 410	-3.5%
TOTAL	All GenBank sequences	451 278 177 138	33.1%

Taxonomic

SYN

PHG

ENV

VRL

BCT

PLN

MAM

VRT

PRI

UNA

ROD

INV

Functional

TSA

WGS

PAT

GSS

EST

HTG

STS

HTC

9

April 2015

PubMed has more than 22 million citations for biomedical literature



PubMed

PubMed comprises more than 22 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.

PubReader

A whole new way to read scientific literature at PubMed Central

Using PubMed

[PubMed Quick Start Guide](#)

[Full Text Articles](#)

[PubMed FAQs](#)

[PubMed Tutorials](#)

[New and Noteworthy](#)

PubMed Tools

[PubMed Mobile](#)

[Single Citation Matcher](#)

[Batch Citation Matcher](#)

[Clinical Queries](#)

[Topic-Specific Queries](#)

More Resources

[MeSH Database](#)

[Journals in NCBI Databases](#)

[Clinical Trials](#)

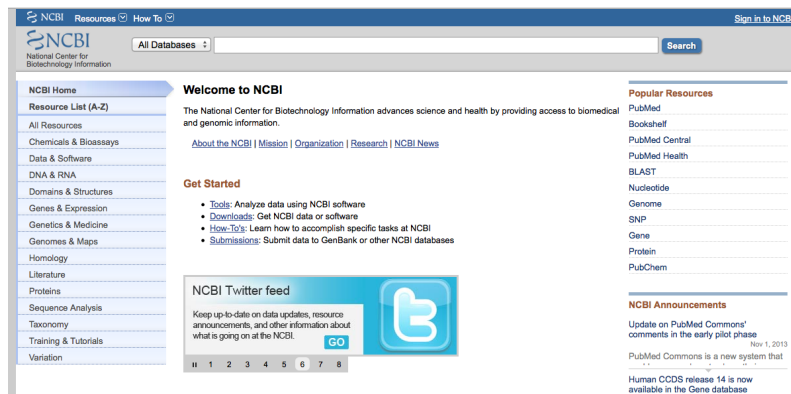
[E-Utilities](#)

[LinkOut](#)

<http://www.ncbi.nlm.nih.gov/pubmed>

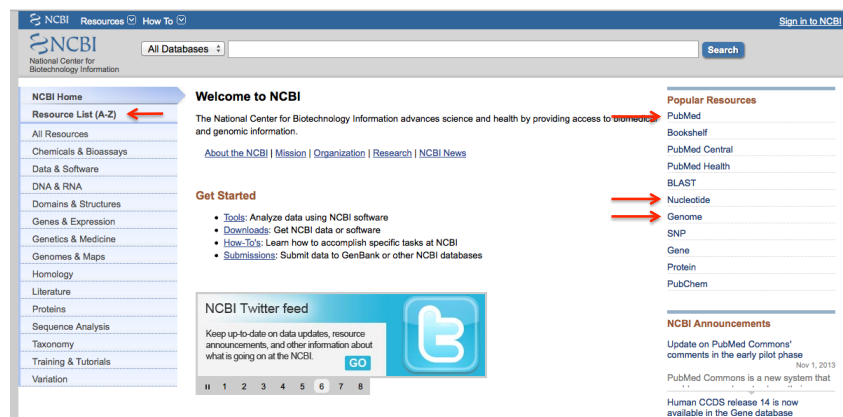
April 2015

The data in GenBank can be accessed via the NCBI Entrez system – a search engine that queries the databases NCBI



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

From the NCBI homepage, many resources can be easily accessed



[illegible]

The data in GenBank can be accessed via the NCBI Entrez system – a search engine that queries the databases at the National Center for Biotechnology Information (NCBI)


NCBI
National Center for
Biotechnology Information

Resources
How To

Sign In to NCBI

All Databases
Bax

Search

NCBI Home
Resource List (A-Z)
All Resources
Chemicals & Biologics
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](#) | [Research](#) | [NCBI News](#)

Get Started

- [Tools](#): Analyze data using NCBI's software
- [Downloads](#): Get NCBI data or software
- [How-To's](#): Learn how to accomplish specific tasks at NCBI
- [Submissions](#): Submit data to GenBank or other NCBI databases

NCBI Twitter feed

Keep up-to-date on data updates, resource announcements, and other information about what is going on at the NCBI.

GO

Popular Resources

- PubMed
- Bookshelf
- PubMed Central
- PubMed Health
- BLAST
- Nucleotide
- Genome
- SNP
- Gene
- Protein
- PubChem

NCBI Announcements

Update on PubMed Commons' comments in the early pilot phase

Nov 1, 2013

PubMed Commons is a new system that

Human CCDB release 14 is now available in the Gene database

The Genome Access Course

GQuery

NCBI Global Cross-database Search

Search NCBI databases

Literature

22286	PubMed: scientific & medical abstracts/citations	19	MeSH: ontology used for PubMed indexing
24966	PubMed Central: full-text journal articles	287	Books: books and reports
41	NLM Catalog: books, journals and more in the NLM Collections	305	Site Search: NCBI web and FTP site index

Health

25	PubMed Health: clinical effectiveness, disease and drug reports	15	ClinVar: human variations of clinical significance
16	MedGen: medical genetics literature and links	98	OMIM: online mendelian inheritance in man
(none)	GTR: genetic testing registry	(none)	OMIA: online mendelian inheritance in animals
(none)	dbGaP: genotype/phenotype interaction studies		

Organisms

(none) Taxonomy: taxonomic classification and nomenclature catalog

Nucleotide Sequences

7144	Nucleotide: DNA and RNA sequences	6	SRA: high-throughput DNA and RNA sequence read archive
7	GSS: genome survey sequences	9	PopSet: sequence sets from phylogenetic and population studies
898	EST: expressed sequence tag sequences	951	Probe: sequence-based probes and primers

15
April 2015

The Genome Access Course

Nucleotide

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

Send to:

Filter your results:

Found 8049 nucleotide sequences. Nucleotide (7144) EST (898) GSS (7)

See **BAX** BCL2-associated X protein in the Gene database
bax in Homo sapiens | Mus musculus | Rattus norvegicus | All 109 Gene records

Results: 1 to 20 of 7144

1. **Rattus norvegicus bax mRNA, partial cds**
Accession: U32098.1 GI: 975869
GenBank FASTA Graphics Related Sequences

2. **Piscirickettsia salmonis alanine racemase (atr) gene, partial cds; and BAX (bax), DNA repair enzyme RadA (radA), 17 kDa antigen (ospA), and transposase (tnpA) genes, complete cds**
4,983 bp linear DNA
Accession: AF184152.1 GI: 10441343
GenBank FASTA Graphics Related Sequences

3. **Danio rerio Bax (bax) gene, complete cds**
579 bp linear DNA
Accession: AF231015.1 GI: 7673631
GenBank FASTA Graphics Related Sequences

All (7144)
Bacteria (2599)
INSDC (GenBank) (4555)
cDNA (4094)
RefSeq (2542)

Top Organisms [Tree]
Escherichia coli (1359)
Saccharomyces cerevisiae (406)
Homo sapiens (375)
synthetic construct (301)
Candida albicans (300)
All other taxa (5305)
More...

Find related data
Database: Select
Find Items

16
April 2015

8

Nucleotide search results for 'Bax'. The first result is highlighted: 4,734,438 bp circular DNA. Accession: NC_016514.1. Other results include Enterobacter cloacae subsp. cloacae ENHKU01 chromosome, complete genome, Homo sapiens BCL2-associated X protein (BAX), transcript variant sigma, mRNA, and Homo sapiens BCL2-associated X protein (BAX), transcript variant alpha, mRNA.

Accession numbers are a string of letters and/or numbers that corresponds to a molecular sequence

Examples (all for beta globin, *HBB*):

<i>X02775</i>	GenBank genomic DNA sequence	DNA
<i>NG_000007.3</i>	RefSeqGene	
<i>rs192792910</i>	dbSNP (single nucleotide polymorphism)	
<i>AA970968.1</i>	An expressed sequence tag	RNA
<i>NM_000518.4</i>	RefSeq DNA sequence (from a transcript)	
<i>NP_000509.1</i>	RefSeq protein	Protein
<i>CAA00182.1</i>	GenBank protein	
<i>Q14473</i>	SwissProt protein	
<i>1YE0 B</i>	Protein Data Bank structure record	

There are several efforts, including the RefSeq project, to compile high quality data from the source datasets

"The goal of the NCBI RefSeq project is to provide an accurate, non-redundant, and comprehensive collection of naturally occurring DNA, RNA, and protein molecules for major organisms."

DNA



RNA



Protein



Function

Table 1. The RefSeq accession number format and molecule types.

Accession prefix	Molecule type	Comment
AC_	Genomic	Complete genomic molecule, alternate assembly
NC_	Genomic	Complete genomic molecule, reference assembly
NG_	Genomic	Incomplete genomic region
NT_	Genomic	Contig or scaffold, clone-based or WGS ^a
NW_	Genomic	Contig or scaffold, primarily WGS ^a
NS_	Genomic	Environmental sequence
NZ_ ^b	Genomic	Unfinished WGS
NM_	mRNA	
NR_	RNA	
XM_ ^c	mRNA	Predicted model
XR_ ^c	RNA	Predicted model
AP_	Protein	Annotated on AC_ alternate assembly
NP_	Protein	
YP_ ^c	Protein	
XP_ ^c	Protein	Predicted model
ZP_ ^c	Protein	Predicted model, annotated on NZ_ genomic records

^a Whole Genome Shotgun sequence data.

^b An ordered collection of WGS for a genome.

^c Computed.



April 2015

The RefSeq Project uses status codes and has extensive additional resources at the NCBI

RefSeq Status codes

Code	Description
GENOME ANNOTATION	The RefSeq record is provided via automated processing and is not subject to individual review or revision between builds.
INFERRED	The RefSeq record has been predicted by genome sequence analysis, but it is not yet supported by experimental evidence. The record may be partially supported by homology data.
PREDICTED	The RefSeq record has not yet been subject to individual review, and some aspect of the RefSeq record is predicted.
PROVISIONAL	The RefSeq record has not yet been subject to individual review. The initial sequence-to-gene name associations have been established by outside collaborators or NCBI staff.
REVIEWED	The RefSeq record has been reviewed by NCBI staff or by a collaborator. The NCBI review process includes assessing available sequence data and the literature. Some RefSeq records may incorporate expanded sequence and annotation information.
VALIDATED	The RefSeq record has undergone an initial review to provide the preferred sequence standard. The record has not yet been subject to final review, at which time additional functional information may be provided.
WGS	The RefSeq record is provided to represent a collection of whole genome shotgun sequences. These records are not subject to individual review or revisions between genome updates.

April 2015

Organism Specific Databases

EcoCyc
A member of the BioCyc database collection

Pathway Tools FBA Tutorial
October 17-18th
Registration Open

LOG IN | Why Log In? | Create New Account

Quick Search | Gene Search

Searching *Escherichia coli* K-12 substr. MG1655 [change organism database](#)

Home Search Tools Help

About EcoCyc
Project Overview
Guide to EcoCyc
Guided Tour
Instructional Videos
Publications
Update History
Steering Committee
Credits
Contact Us

ECOCYC OVERVIEW
EcoCyc is a scientific database for the bacterium *Escherichia coli* K-12 MG1655. The EcoCyc project performs literature-based curation of the entire genome, and of transcriptional regulation, transporters, and metabolic pathways. [\[project overview\]](#)

New to EcoCyc? Take the [guided tour](#) of the EcoCyc.org Web site, watch our [free online instructional videos](#), or read our 2011 article: ["EcoCyc: a comprehensive database of *Escherichia coli* biology."](#)

Need help? Please [contact us](#).

ECOCYC NEWS

- The first *E. coli* publication

The oldest *E. coli* paper we've found indexed in PubMed is [The Importance to the Surgeon of Familiarity with the Bacillus Coli Communis](#) by Roswell Park. The oldest paper around is the original by Escherich. It's not indexed in PubMed, but you can click the following link to read [Die Darmbakterien des Neugeborenen und Säuglings](#).

Information
Gene Cross-References
Unsequenced Exons
Conferences

TB Database
AN INTEGRATED PLATFORM FOR TUBERCULOSIS RESEARCH

QUICK SEARCH

Home Publications Expression Data **Genomic Data** Systems Biology Proteins Advanced Search Help

Genomes Genes BLAST Diversity Sequencing Comparative Analysis Feature Search Genome Browser Operon Browser Download

TB Genomes Database

Genomes
Genes
BLAST
Diversity Sequencing
Comparative Analysis
Feature Search
Systems Biology
Genome Browser
Operon Browser
Publications

Yak

21
April 2015

Organism Specific Databases

UC DAVIS
KOMP Repository
KNOCKOUT MOUSE PROJECT

Advanced Gene Searches
Gene Ontology
Disease
Microarray
Literature

Contact Information

Home MyKOMP Log In New Account Catalog

KOMP2 mice now available from the KOMP Repository: [click here](#)

Search gene or ID

You can search by [gene name](#), [gene ID](#) or [project](#)

Products...
Gene symbol...

Browse

Home Search Tools Downloads Links About

Welcome to The Yak Genome Database

The recent sequenced yak genome Assembly Yak_1.1 has a size of 2,657 Mb with 65X coverage. A total of 22,282 protein-coding genes were predicted based on RNA-seq, homology and *ab initio* gene prediction. The high coverage permitted identification of 2.2 million heterozygous single nucleotide variants (SNVs) within the sequenced individual.

YGD includes GBrowse genome browsers, BLAST databases and other tools to facilitate the study of gene function and genome evolution.

Recent News

22
April 2015

Worked examples 1: NCBI gene information

April 2015

ACCTTCACCTTGAGCTAGGAAGGGCGCATTACCTTACTGGAGGTAAAGTAAGTTTGGCCTCTGTTG
 TCATGTGAGTCGTATCTAGAAACGCAGCAAGGGTGTGGGAATATCTTAACACAGGGTCTCATTGG
 AGCCCGTAGTTTCACAGAAACCTCAGGCGGTGCGCATTACGGCTGTGACATTTGAGGTGAAGGTGT
 CAGCATATATCCACCCCCACCCCAACCCCTGGCTACTAAAGTAACATTGGTTTGTGTTTGTGAG
 ACAGGATCTTGGTGTGTAGTTGAGGCTGGCCTCAAAGCCTAGGGTGGCCTGGGACAAAGTCAGCC
 TTCTGAGTGTGTTGGG/ ACCCTTAGTACTAA
 GCACATGTGTATGC/ ACTATCAAATACTG
 TATGACCTGTTGGA/ GGCTGGGGTTATCA
 CAGAGAAAGGAGCT/ GGTATTTTCTCAAT
 TACTGATCAAGGGG/ TAGTCTTGGGTTCT
 ATAAGAGAGGAGGC/ CTCTGCATCAGCTC
 CTGCTTCCTGACCT/ AGCAATGTGAAAAG
 TGTAAAGCCGAATA/ TGCAGGAATAGAAA
 CCCTGACTAAGACAGAGCTGTGCTTTTTATAGTGAACACTTGAAAAATAACTCAGTCTAGAGAA
 TTTGTTGGAAATGTTTACAATCAAACCTGTAGCAACATTGATAGATGTATTGTTTAACTGAATATA
 TAAACAGGCATTTTAAATGTAGAATTTCAAGCTTCATAATAGATAAAATTAGCCTCATCTAATTT
 TATTAGGCAGCTATTCTTTTAAAAATTAAGGCAGTTTGGGTAAGGGATATGAGACGTCAAGATGTA
 TAAAAAACCACAAAGATTATCACTTAAGATTTTACTGAAAAGATGCTGCAGGTATGTGCAGTCT
 CATGTTTGGCAAGTATGAGGTGCACTAGTCCCAACTGTCCTGAAATTCACGTGTAGAATGGCCTGA
 CCATGAGCTTTAAGAAGGTCTGTGTGTTTCTGCCTCCCAAGTACTGTGCAAGACTTGCATCAGCA
 TTCCTGGTCTTAGTAGAATGCTGTGTGTTTGGTGTGGAACATGGCCCCAAGGGCAGATTTAA
 TTTCTGTTCTTTGTACTATGTGATGTTTCTAAACATT

More advanced:
Gene expression
- RNA

24

April 2015

There are three major (and many minor) gene expression resources: UniGene, GEO, and BioGPS

- DNA

↓

RNA

↓

Protein

↓

Function

- *UniGene*: UniGene computationally identifies transcripts from the same locus; analyzes expression by tissue, age, and health status; and reports related proteins (protEST) and clone resources.

<http://www.ncbi.nlm.nih.gov/sites/entrez?db=unigene> *
 - *GEO profiles*: stores individual gene expression profiles from curated DataSets in the Gene Expression Omnibus (GEO) repository. Search for specific profiles of interest based on gene annotation or pre-computed profile characteristics.

<http://www.ncbi.nlm.nih.gov/geo/>
 - *BioGPS*: a free extensible and customizable gene annotation portal, a complete resource for learning about gene and protein function.

<http://biogps.gnf.org/>

April 2015

BioGPS is a very useful stop to ask questions about where a gene is expressed and it is now customizable



A free *extensible and customizable* **gene annotation portal**, a complete resource for learning about **gene and protein function**.



Plugins » Datasets »

Simple to use

- 1 **Search**
for your gene of interest
- 2 **View**
the gene annotation report
- 3 **Browse**
the gene report layouts
- 4 **Build**
your own gene report

[Details »](#)

Search genes here:

Press Ctrl+Enter or click **Search**

Example Searches
(click to try these examples)

- Gene Symbol(s)
- Wildcard queries
- Gene Ontology
- Affymetrix IDs
- Interpro
- Genomic interval

User Love

News and Musings

26

April 2015

UniProt combines the UniProtKB, UniRef, and UniParc databases in a comprehensive resource

<http://www.uniprot.org>



UniProtKB = SWISS-PROT (manually annotated)+
TrEMBL (automatically annotated and NOT reviewed)

UniRef = sequence clusters

UniParc = Sequence archive

```

>sp|P31273|HXC8_HUMAN Homeobox protein Hox-C8 (Hox-3A) - Homo sapiens (Human).
MSSYFVNPLFSKYKAGESLEPAYDCRFPSVGRSHALVYGPQGSAPGFQHASHHVQDFF
HHGTSGISNSGYQQNPCLSCHGDASKFYGYEALPRQSLYGAQQEASVVQYPDCKSSANT
NSSEGQGHNLQNSSPSLMFPWMRPHAPGRRSGRQTSRYQTLELEKEFLFNPYLTRKRI
EVSHALGLTERQVKIWFQNRMRKWKKENNKDKLPGARDEEKVEEEGNEEEEEEEEEKEEN
KD
  
```

April 2015

PDB

The screenshot shows the PDB website interface. At the top, there's a navigation bar with 'PDB' and 'Protein Data Bank' logos. Below this is a search bar with a dropdown menu showing 'Everything', 'Author', 'Macromolecule', 'Sequence', and 'Ligand'. The main content area is titled 'Biological Macromolecular Resource' and features a 'Molecule of the Month' section for 'SNARE Proteins'. There are also links to 'Learn: Featured Molecules', 'Protein Synthesis', and 'Drug and Drug Target Mapping'. The interface is clean and professional, with a blue and white color scheme.

28
April 2015

The ExPASy Portal provides databases and software tools for proteomics, genomics, phylogeny, systems biology, population genetics, and transcriptomics

DNA

↓

RNA

↓

Protein

↓

Function

April 2015

Worked examples 2: Gene Expression

April 2015

A. WE1

NCBI: EntrezGene - Find all functional information for human *BAX*
(or your gene of interest)

The screenshot shows the NCBI homepage. In the top navigation bar, there is a search bar with the text 'Gene' and a dropdown menu. The dropdown menu is open, showing a list of options: All Databases, NCBI Web Site, Databases, PubMed, Protein, Nucleotide, EST, CSS, Structure, Genome, Books, Cancer/Chromosomes, Conserved Domains, 3D Domains, Gene, Genome Project, dbGaP, GENSAT, GEO Profiles, and GEO Datasets. The 'Gene' option is highlighted. To the right of the search bar, there is a search button labeled 'Search'. Below the search bar, there is a section titled 'Resources' with a list of links: NCBI Home, All Resources (A-Z), Literature, DNA & RNA, Proteins, Sequence Analysis, Genes & Expression, Genomes, Maps & Markers, Domains & Structures, Genetics & Medicine, Taxonomy, Data & Software, Training & Tutorials, Homology, Small Molecules, and Variation. In the center of the page, there is a section titled 'PubMed Central' with a link to 'Free Full Text. Over 1,500,000 articles from over 450 journals. Linked to PubMed and fully searchable.' Below this, there is a 'How To...' section with a list of links: Obtain the full text of an article, Retrieve all sequences for an organism or taxon, Find a homolog for a gene in another organism, Find genes associated with a phenotype or disease, Design PCR primers and check them for specificity, Find the function of a gene or gene product, and Find syntenic regions between the genomes of two organisms. To the right of the search bar, there is a 'Help Us Improve' section with a link to 'Send us feedback on the new home page and guide pages design'. Below this, there is a section titled 'NLM/NCBI H1N1 Flu Resources' with a link to 'FLU.GOV' and a 'VISIT FLU.GOV' button. Below the 'FLU.GOV' button, there is a list of links: Newest H1N1 influenza sequences, Submit flu sequences to GenBank, Latest H1N1 citations in PubMed, MedlinePlus (consumer health information), and Enviro-Health links.

1. NCBI home page: drop down menu and highlight 'Gene'.

Type in 'BAX' and 'Go'.

Name/Gene ID	Description	Location	Aliases
☐ BAX ID: 581	BCL2-associated X protein [<i>Homo sapiens</i> (human)]	Chromosome 19, NC_000019.10 (48954860..48961798)	BCL2L4
☐ Bax ID: 12028	BCL2-associated X protein [<i>Mus musculus</i> (house mouse)]	Chromosome 7, NC_000073.6 (45461695..45466899, complement)	
☐ Bax ID: 24887	Bcl2-associated X protein [<i>Rattus norvegicus</i> (Norway rat)]	Chromosome 1, NC_005100.4 (101451801..101457207, complement)	
☐ bax ID: 12933551	hypothetical protein [<i>Escherichia coli</i> str. K-12 substr. W3110]	NC_007779.1 (3903238..3904062)	Y75_p3605
☐ BAX ID: 280730	BCL2-associated X protein [<i>Bos taurus</i> (cattle)]	Chromosome 18, AC_000175.1 (55985166..55989379)	BOS_17596

2. BAX in different species, choose human

Entrez gene entry for Human *BAX* Gene - scroll down to find a wealth of information

BAX BCL2-associated X protein [*Homo sapiens* (human)]

Gene ID: 581, updated on 5-Oct-2014

Summary

Official Symbol BAX provided by [HGNC](#)

Official Full Name BCL2-associated X protein provided by [HGNC](#)

Primary source [HGNC:HGNC:959](#)

See related [Ensembl:ENSG00000087088](#); [HPRD:02498](#); [MIM:600040](#); [Vega:OTTHUMG00000160476](#)

Gene type protein coding

RefSeq status REVIEWED

Organism [Homo sapiens](#)

Lineage Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Haplorhini; Catarrhini; Hominidae; Homo

Also known as BCL2L4

Summary The protein encoded by this gene belongs to the BCL2 protein family. BCL2 family members hetero- or homodimers and act as anti- or pro-apoptotic regulators that are involved in a wide cellular activities. This protein forms a heterodimer with BCL2, and functions as an apoptotic. This protein is reported to interact with, and increase the opening of, the mitochondrial voltage dependent anion channel (VDAC), which leads to the loss in membrane potential and the release of cytochrome c from mitochondria.

Table of contents

- Summary
- Genomic context
- Genomic regions, transcripts, and products
- Bibliography
- Phenotypes
- Variation
- HIV-1 interactions
- Pathways from BioSystems
- Interactions
- General gene information
- Homology, Gene Ontology
- Protein information
- Sequence Sequences (RefSeq)
- References
- Links

Genomic regions, transcripts, and products

Go to [reference sequence details](#)

Genomic Sequence: [NC_000019.10 chromosome 19 reference GRCh38 Primary Assembly](#)

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

NC_000019.10: 49M..49M (9.0Kbp)

Genes, NCBI Homo sapiens Annotation Release 106

CCDS Feat

CCDS127

CCDS127

CCDS127

CCDS127

Genes, En

560000008708

Bibliography

Related articles in PubMed

1. [Foci of lymphocytotoxicity in human B-cell lymphoma] Shvets VN. Med Res Rev. 1998;18(1):1-10.
2. NMR studies of the BAX protein. Yao S, et al. Arch Biochem Biophys. 2000;381(1):1-10.
3. ROS-mediated JNK activation in B cells. Park GB, et al. Int J Cancer. 2000;88(1):1-10.
4. BaxΔ2 Family A... Haferkamp B, et al. J Biol Chem. 2000;275(1):1-10.
5. Lack of function...

GeneRIFs: Gene References Into Functions [What's a GeneRIF?](#)

1. miR-511 regulates the growth of radioresistant lung cancer cells by increasing BAX expression through TRIB2.
2. Superoxide activates mTOR-eIF4E-Bax route to induce enhanced apoptosis in leukemic cells.
3. SRF can induce the apoptosis of EBV-transformed B cells through JNK/p38-MAPK activation. SRF triggered the generation of reactive oxygen species (ROS), translocation of Bax into the mitochondria, disruption of mitochondrial membrane potential.
4. a missense mutation in the C-terminal transmembrane domain of proapoptotic BAX (G179E) was found, which abrogated BAX anchoring to mitochondria and blocked ABT-199-induced apoptosis both in vitro and in vivo
5. This meta-analysis suggests that the Bax-249C>A polymorphism is not an independent prognostic factor for overall survival in patients with gastric cancer.

mRNA and Protein(s)

[NM_001291428.1](#) → [NP_001278357.1](#) apoptosis regulator BAX isoform 1

[See proteins identical to NP_001278357.1](#)

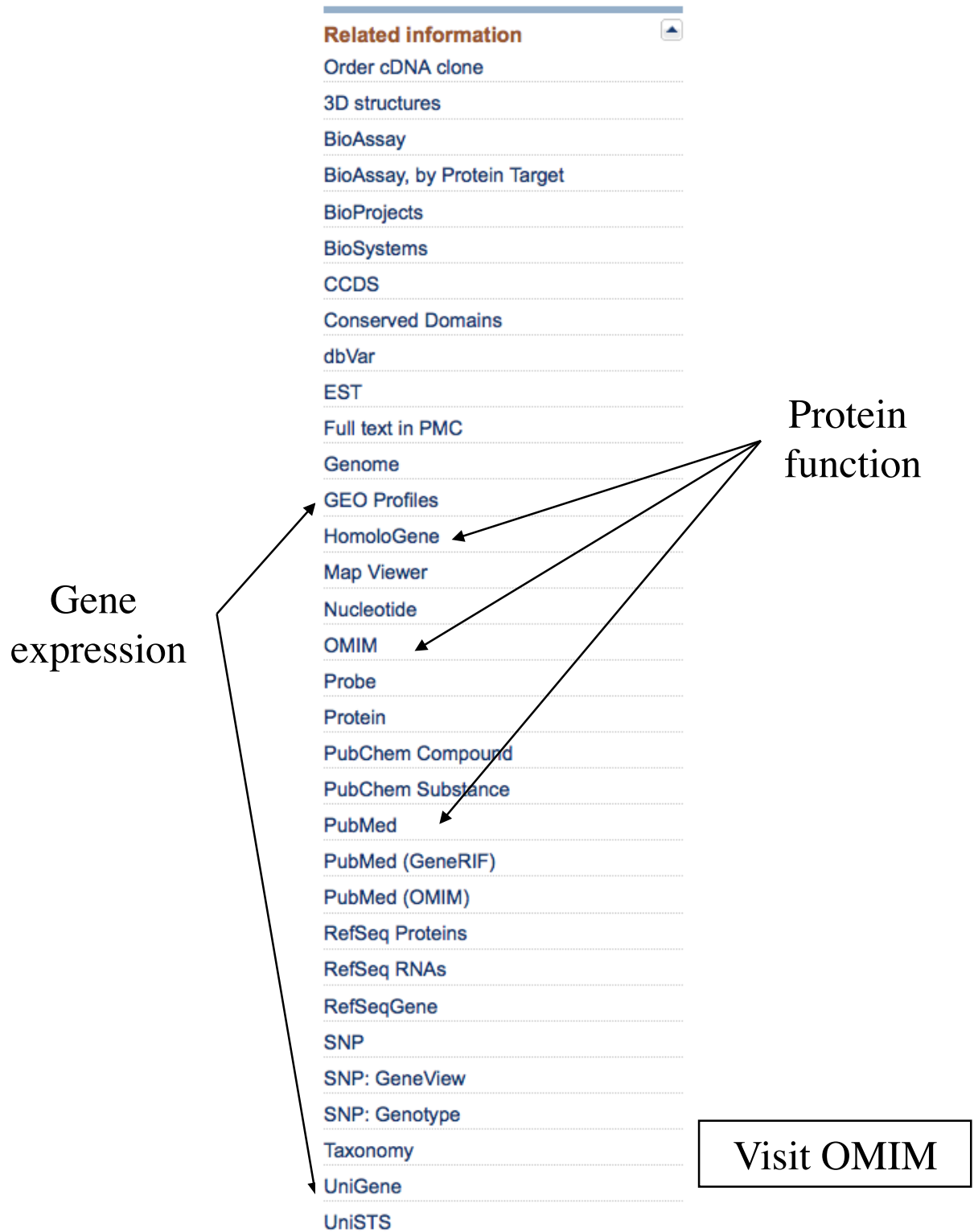
Status: REVIEWED

Description	Transcript Variant: This variant (1) encodes the longest isoform (1).
Source sequence(s)	AI565203 , AK291076 , BE396495 , BM673184 , CN281003
UniProtKB/Swiss-Prot	Q07812

Conserved Domains (2) [summary](#)

cd06845	Bcl-2_like; Apoptosis regulator proteins of the Bcl-2 family, named after B-cell lymphoma 2. This alignment model spans what have been described as Bcl-2 homology regions BH1, BH2, BH3, and BH4. Many members of this family have an additional C-terminal transmembrane ...
Location:25 → 158	
TIGR00865	bcl-2; apoptosis regulator
Location:5 → 182	

Links to other Entrez databases and other ancillary databases



[Home](#) | [About](#) | [Statistics](#) | [Downloads](#) | [Help](#) | [External Links](#) | [Copyright](#) | [Contact Us](#)

Select Language

Sort by: ☒ Relevance ☐ Date updated

Advanced Search: [OMIM](#), [Clinical Synopses](#), [OMIM Gene Map](#)
 Search History: [View](#), [Clear](#)

***600040**

BCL2-ASSOCIATED X PROTEIN; BAX

*HGNC Approved Gene Symbol: **BAX***

*Cytogenetic location: **19q13.33** Genomic coordinates (GRCh37): **19:49,458,116 - 49,465,054*** (from NCBI)

Gene Phenotype Relationships

Location	Phenotype	Phenotype MIM number
19q13.33	Colorectal cancer	
	T-cell acute lymphoblastic leukemia	

TEXT

Description

The proapoptotic BAX protein induces cell death by acting on mitochondria.

Cloning

[Oltvai et al. \(1993\)](#) identified BAX as a protein partner of BCL2 ([151430](#)).

Gene Function

Development as well as maintenance of many adult tissues is achieved by several dynamically regulated processes that include cell proliferation, differentiation, and programmed cell death. [Oltvai et al. \(1993\)](#) noted that, in the latter process, cells are eliminated by a highly characteristic suicide program called apoptosis. The best-defined genetic pathway of cell death exists in the nematode *Caenorhabditis elegans*. Two autosomal recessive death effector genes, *ced-3* and *ced-4*, are required for the death of all 131 cells destined to die during worm development. One autosomal dominant death repressor gene, *ced-9*, can save those cells in its gain-of-function form. This implies that both effector and repressor genes also exist within each mammalian cell death pathway. BCL2 is one such mammalian gene that has been identified; it functions as a repressor of programmed cell death.

[Table of Contents - *600040](#)
 External Links:
[Genome](#)
[DNA](#)
[Protein](#)
[Gene Info](#)
[Variation](#)
[Animal Models](#)
[Cellular Pathways](#)

Summary of information for human BAX gene

- includes any relevant disease association
- check for ‘last update’ at the bottom

Click back to EntrezGene entry for *BAX*
Click 'homologene'

HomoloGene:7242. Gene conserved in Euteleostomi

Genes
Genes identified as putative homologs of one another of the construction of HomoloGene.

List of all 'catalogued' orthologues at NCBI. Click ID 7242

 BAX, <i>H.sapiens</i> BCL2-associated X protein	 NP_620116.1 192 aa
 BAX, <i>P.troglodytes</i> BCL2-associated X protein	 XP_001154 192 aa
 BAX, <i>M.mulatta</i> BCL2-associated X protein	 XP_001112 192 aa
 BAX, <i>C.lupus</i> BCL2-associated X protein	 NP_00100 192 aa
 BAX, <i>B.taurus</i> BCL2-associated X protein	 NP_77631 192 aa
 Bax, <i>M.musculus</i> BCL2-associated X protein	 NP_03155 192 aa
 Bax, <i>R.norvegicus</i> Bcl2-associated X protein	 NP_058755.2 192 aa
 bax, <i>X.tropicalis</i> BCL2-associated X protein	 NP_989185.1 221 aa
 baxa, <i>D.rerio</i> bcl2-associated X protein, a	 NP_571637.1 192 aa

View functional information for mouse *Bax*. Go back to Homologene and click 'M. musculus *Bax*.'

Display Settings: ☒ Full Report

Bax BCL2-associated X protein [*Mus musculus*]
Gene ID: 12028, updated on 4-Nov-2010

Summary

Official Symbol [Bax](#) provided by [MGI](#)
Official Full Name [BCL2-associated X protein](#) provided by [MGI](#)
Primary source [MGI:99702](#)
See related [Ensembl:ENSMUSG00000003873](#)
Gene type protein coding
RefSeq status VALIDATED
Organism [Mus musculus](#)
Lineage Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Muroidea; Muridae; Murinae; Mus; Mus
Also known as Bax

Entrez Gene for Mouse *Bax*. Scroll to find 'Links to external resources' and click on MGI

Links to other resources

[MGI](#)

[Ensembl](#)

[Evidence Viewer](#)

[Gene Expression Database \(GXD\) at MGI](#)

[UCSC](#)

[MGC](#)

[KEGG](#)

[Reactome](#)


[Celebrate 25 years with us](#)

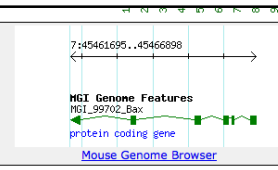
[About](#)
[Help](#)
[FAQ](#)

[Home](#)
[Genes](#)
[Phenotypes](#)
[Human Disease](#)
[Expression](#)

[Search](#)
[Download](#)
[More Resources](#)
[Submit Data](#)
[Find Mice \(IMSR\)](#)
[Analysis Tools](#)

Bax
Gene Detail

Symbol	Bax
Name	BCL2-associated X protein
ID	MGI:99702
Feature Type	protein coding gene
Genetic Map	Chromosome 7 29.32 cM, cytoband B5 Detailed Genetic Map ± 1 cM Mapping data(11)
Sequence Map	Chr7:45461695-45466898 bp, - strand From NCBI annotation of GRCm38 Get FASTA 5204 bp ± 0 kb flank VEGA Genome Browser Ensembl Genome Browser UCSC Browser NCBI Map Viewer
Vertebrate homology	HomoloGene:7242 Vertebrate Homology Class 1 human; 1 mouse; 1 rat; 1 chimpanzee; 1 rhesus macaque; 1 cattle; 1 dog; 1 western clawed frog; 1 zebrafish Protein SuperFamily: apoptosis regulator BAX Gene Tree: Bax
Human homologs	Human Homolog BAX, Bcl2-associated X protein NCBI Gene ID 581 neXtProt AC NX_Q07812 Human Synonyms BCL2L4 Human Chr (Location) 19q12-13.4; chr19:48954860-48961798 (+) <i>GRCh38</i>
Mutations, alleles, and phenotypes	All mutations/alleles(41) : Gene trapped(35) Targeted(6) Homozygous null mutants display hyperplasia of thymocytes and B cells, reproductive failure with abnormal germ cells and gonadal morphology, and reduced cell death in the CNS and PNS. Female mutants exhibit a prolonged ovarian lifespan.



MGI Genome Features
MGI_99702_Bax
protein coding gene
[Mouse Genome Browser](#)

Information for *Bax* in Organism Specific Databases, e.g Mouse Genome Informatics (MGI).

To view disease relevant information, click ‘all phenotypic alleles’.

Export: [Text File](#) [Excel File](#)

Allele Symbol Gene; Allele Name	Chr	Synonyms	Category	Abnormal Phen Systems
Bax^{tm1Dlo} BCL2-associated X protein; targeted mutation 1, Dennis Y Loh	7	bax ⁻	Targeted (Null/knockout)	cellular, nervous
Bax^{tm1Ofxa} BCL2-associated X protein; targeted mutation 1, Osborne FX Almeida	7	bax ⁻	Targeted (Null/knockout)	cellular, nervous
Bax^{tm1Sik} BCL2-associated X protein; targeted mutation 1, Stanley J Korsmeyer	7	Bax ⁻ , Bax1	Targeted (Null/knockout)	cellular, endocrine, hematopoietic, immune system, renal/urinary, reproductive, respiratory, skeleton, tumorigenesis
Bax^{tm2Sik} BCL2-associated X protein; targeted mutation 2, Stanley J Korsmeyer	7	Bax ^{fl}	Targeted (Conditional ready, No functional change)	hematopoietic, immune, mortality/aging, renal/urinary, skeleton, tumorigenesis

All strains of mouse with mutations in *Bax*.

Click on strain name to reveal phenotypic information for individual strains.

Targeted Mouse Detail																			
Nomenclature Mutation origin Mutation description Phenotypes Find Mice (IMSR) Notes References																			
Nomenclature	Symbol:	Bax ^{tm1Dlo}																	
	Name:	BCL2-associated X protein; targeted mutation 1, Dennis Y Loh																	
	MGI ID:	MGI:2158306																	
	Synonyms:	bax-																	
Mutation origin	Gene:	Bax Location: Chr7:45461695-45466898 bp, - strand Genetic Position: Chr7, 29.32 cM, cytoband B5																	
	Germline Transmission:	Earliest citation of germline transmission: J:39573																	
	Parent Cell Line:	E14 (ES Cell)																	
Mutation description	Strain of Origin:	129P2/OlaHsd																	
	Allele Type:	Targeted (Null/knockout)																	
	Mutations:	Insertion, Intragenic deletion																	
Phenotypes	A 5.6 kb genomic fragment containing exons 2-6 was replaced with a neomycin selection cassette. (J:39573)																		
	Key:	hm homozygous ht heterozygous tg involves transgenes ✓ phenotype observed cn conditional genotype cx complex: > 1 genome feature ot other: hemizygous, indeterminate, ... N normal phenotype																	
	Genotypes:	<table><thead><tr><th>Genotype</th><th>Allelic Composition</th><th>Genetic Background</th><th>Cell Line(s)</th></tr></thead><tbody><tr><td>hml</td><td>Bax^{tm1Dlo}/Bax^{tm1Dlo}</td><td>Involves: 129P2/OlaHsd</td><td></td></tr><tr><td>cx2</td><td>Bax^{tm1Dlo}/Bcl2^{tm1Dlo}</td><td>Involves: 129P2/OlaHsd</td><td></td></tr></tbody></table>			Genotype	Allelic Composition	Genetic Background	Cell Line(s)	hml	Bax ^{tm1Dlo} /Bax ^{tm1Dlo}	Involves: 129P2/OlaHsd		cx2	Bax ^{tm1Dlo} /Bcl2 ^{tm1Dlo}	Involves: 129P2/OlaHsd				
Genotype	Allelic Composition	Genetic Background	Cell Line(s)																
hml	Bax ^{tm1Dlo} /Bax ^{tm1Dlo}	Involves: 129P2/OlaHsd																	
cx2	Bax ^{tm1Dlo} /Bcl2 ^{tm1Dlo}	Involves: 129P2/OlaHsd																	
Phenotypes:	<table><thead><tr><th>Affected Systems</th><th>hml</th><th>cx2</th></tr></thead><tbody><tr><td>show or hide all annotated terms</td><td></td><td></td></tr><tr><td>cellular</td><td>✓</td><td>✓</td></tr><tr><td>mortality/aging</td><td>✓</td><td>✓</td></tr><tr><td>nervous system</td><td>✓</td><td>✓</td></tr></tbody></table>				Affected Systems	hml	cx2	show or hide all annotated terms			cellular	✓	✓	mortality/aging	✓	✓	nervous system	✓	✓
	Affected Systems	hml	cx2																
	show or hide all annotated terms																		
	cellular	✓	✓																
mortality/aging	✓	✓																	
nervous system	✓	✓																	
View phenotypes for all genotypes (concatenated display).																			

Click on arrow:

Decrease apoptosis - not surprising for a

Click on arrow:
Decrease apoptosis - not surprising for a proapoptotic gene!

cellular	decreased apoptosis (J:66953)
	decreased neuron apoptosis (J:39573 , J:66953)
nervous system	decreased neuron apoptosis (J:39573 , J:66953)

- E12 brainstem and ventral spinal cord have fewer TUNEL positive cells than normal while E12 primary telencephalic cultures show wildtype levels of apoptosis at 48 hours (J:39573)
- E12.5 primary telencephalic cultures show a small resistance to AraC induced neuronal cell death relative to wild-type (J:66953)

	
Home	Genes / Markers / Clones
Research	General Information
Expression	ZIRC
Antibodies	BLAST
Mutants / Morphants / Tg	Anatomy
Maps	Publications
ZFIN ID: ZDB-GENE-000511-6	
Gene Name: <i>bcl2-associated X protein, a</i>	
Gene Symbol: <i>baxa</i>	
Previous Names: wu:f16e01(1), zBax1(1), bax, f16e01, wu:fc50b10	
Nomenclature History	
GENE EXPRESSION: (current status)	
All expression data: 20 figure(s) from 15 publications	
Directly submitted expression data: 1 figure(s) (1 images) from Thisse et al., 2004 [MGC:66309]	
Wild Type Stages, Structures: Zygote: 1-cell (0.00h-0.75h) to Adult (90d-730d, breeding adult)	
Curated microarray expression: GEO (1)	
MUTANTS AND TARGETED KNOCKDOWNS:	
Phenotype: (current status)	
Knockdown reagents: MO1-baxa (1), MO2-baxa (1)	

Click back to Homologene,
Choose Danio rerio 'bax'
Organism specific database is 'Zfin'

Explore ZFIN to find out more about the BAX gene

Click back to EntrezGene entry for human *BAX* and scroll down to view Gene Ontology terms

Function	Evidence Code	Pubs
BH3 domain binding	IDA	
BH3 domain binding	IPI	PubMed
channel activity	IDA	PubMed
identical protein binding	IPI	PubMed

lipid binding	Items 1 - 25 of 82 << First < Prev Page 1 of 4 Next > Last >>		
protein binding	Process	Evidence Code	Pubs
protein binding	B cell apoptotic process	IDA	PubMed
protein binding	B cell homeostasis	IEA	
protein binding	B cell homeostatic proliferation	IEA	
protein binding	B cell negative selection	IEA	
protein binding	Sertoli cell proliferation	IEA	
protein binding	T cell homeostatic proliferation	IEA	

activation of c	Component	Evidence Code	Pubs
activation of c	Bcl-2 family protein complex	IDA	
activation of c	cytosol	IDA	PubMed
activation of c	cytosol	TAS	
cytochrome c	cytosolic part	IEA	
	endoplasmic reticulum	IDA	PubMed
	endoplasmic reticulum membrane	IDA	PubMed
	mitochondrial outer membrane	TAS	

Three different classifications for Gene Ontology:

- Molecular function:** Activities, such as catalytic or binding activities, that occur at the molecular level.
- Biological process:** A series of events accomplished by one or more ordered assemblies of molecular functions.
- Cellular component:** A component of a cell, but with the proviso that it is part of some larger object; this may be an anatomical structure (e.g. rough endoplasmic reticulum or nucleus) or a gene product group (e.g. ribosome, proteasome or a protein dimer).

See <http://www.geneontology.org/GO.doc.shtml> for more details.

Three different classifications for Gene Ontology:

- Molecular function:** Activities, such as catalytic or binding activities, that occur at the molecular level.
- Biological process:** A series of events accomplished by one or more ordered assemblies of molecular functions.
- Cellular component:** A component of a cell, but with the proviso that it is part of some larger object; this may be an anatomical structure (e.g. rough endoplasmic reticulum or nucleus) or a gene product group (e.g. ribosome, proteasome or a protein dimer).

See <http://www.geneontology.org/GO.doc.shtml> for more details.

Three different classifications for Gene Ontology:

- Molecular function:** Activities, such as catalytic or binding activities, that occur at the molecular level.
- Biological process:** A series of events accomplished by one or more ordered assemblies of molecular functions.
- Cellular component:** A component of a cell, but with the proviso that it is part of some larger object; this may be an anatomical structure (e.g. rough endoplasmic reticulum or nucleus) or a gene product group (e.g. ribosome, proteasome or a protein dimer).

See <http://www.geneontology.org/GO.doc.shtml> for more details.

Three different classifications for Gene Ontology:

- Molecular function:** Activities, such as catalytic or binding activities, that occur at the molecular level.
- Biological process:** A series of events accomplished by one or more ordered assemblies of molecular functions.
- Cellular component:** A component of a cell, but with the proviso that it is part of some larger object; this may be an anatomical structure (e.g. rough endoplasmic reticulum or nucleus) or a gene product group (e.g. ribosome, proteasome or a protein dimer).

See <http://www.geneontology.org/GO.doc.shtml> for more details.

Three different classifications for Gene Ontology:

- Molecular function:** Activities, such as catalytic or binding activities, that occur at the molecular level.
- Biological process:** A series of events accomplished by one or more ordered assemblies of molecular functions.
- Cellular component:** A component of a cell, but with the proviso that it is part of some larger object; this may be an anatomical structure (e.g. rough endoplasmic reticulum or nucleus) or a gene product group (e.g. ribosome, proteasome or a protein dimer).

See <http://www.geneontology.org/GO.doc.shtml> for more details.

Click on 'BH3 domain binding' to view associated GO information.

BH3 domain binding

Term Information

Accession GO:0051434
Name BH3 domain binding
Ontology molecular_function
Synonyms None
Definition Interacting selectively and non-covalently with the BH3 domain of a protein of the Bcl-2 family. The BH3 domain is a potent death domain and has an important role in protein-protein interactions and in cell death. *Source:* [Prosite:PS01259](#), [PMID:9020082](#), [PMID:12133724](#), [PMID:11048732](#), [PMID:9704409](#)
Comment None
History See term [history for GO:0051434](#) at QuickGO
Subset None
Community Unavailable
Related [Link](#) to all **genes and gene products** associated to BH3 domain binding.
[Link](#) to all direct and indirect **annotations** to BH3 domain binding.
[Link](#) to all direct and indirect **annotations download** (limited to first 10,000) for BH3 domain binding.

View other genes with this domain

Associations [Graph Views](#) [Inferred Tree View](#) [Ancestors and Children](#) [Mappings](#)

Free-text filtering ☒

+ document_category: annotation
+ regulates_closure: GO:0051434

No current user filters.

Assigned by
Ontology (aspect)
Evidence type
PANTHER family
Qualifier
Taxon
Direct annotation
Inferred annotation
Annotation extension

Found entities
Total: 58; showing 1-10 Results count (10)

Gene/product	Gene/product name	Qualifier	Direct annotation	Annotation extension	Source	Taxon	Evidence	Evidence with	PANTHER family	Isoform	Reference
Bax	BCL2-associated X protein	BH3 domain binding			MGJ	Mus musculus	ISO	UniProtKB-Q07812	bcl-2 related pthr11256	UniProtKB-Q07813	MGJ:MGJ:4834177
Bcl2l1	Bcl2-like 1	BH3 domain binding			RGD	Rattus norvegicus	ISO	RGD:735516	bcl-2 related pthr11256		RGD:1624291
Mcl1	myeloid cell leukemia sequence 1	BH3 domain binding			RGD	Rattus norvegicus	ISO	RGD:735512			RGD:1624291
Mcl1	myeloid cell leukemia sequence 1	BH3 domain binding			RGD	Rattus norvegicus	ISO	RGD:735513			RGD:1624291
Bcl2l1	BCL2-like 1	BH3 domain binding			MGJ	Mus musculus	ISO	UniProtKB-Q07817	bcl-2 related pthr11256	VEGA-OTTMUSP00000016850	MGJ:MGJ:4834177
Bcl2	B cell leukemia/lymphoma 2	BH3 domain binding			MGJ	Mus musculus	ISO	UniProtKB-P10415	bcl-2 related pthr11256	VEGA-OTTMUSP00000024915	MGJ:MGJ:4834177
Bcl2l1-ps1	Protein Bcl2l1-ps1	BH3 domain binding			UniProtKB	Rattus norvegicus	IEA	Ensembl:ENSP00000302564	bcl-2 related pthr11256		GO_REF:0000019
Bcl2l1-ps1	Bcl2-like 1, pseudogene 1	BH3 domain binding			RGD	Rattus norvegicus	IEA	Ensembl:ENSP00000302564			RGD:1600115
Bcl2	Apoptosis regulator Bcl-2	BH3 domain binding			UniProtKB	Rattus norvegicus	IEA	Ensembl:ENSP00000381185	bcl-2 related pthr11256		GO_REF:0000019
Bax	Apoptosis regulator BAX	BH3 domain binding			UniProtKB	Rattus norvegicus	ISS	UniProtKB-Q07812			GO_REF:0000024

Gene expression: Click back to human BAX entry in Entrez Gene and link to 'Unigene'

1: [BCL2-associated X protein](#)
BAX, *Homo sapiens*
Hs.624291: 139 sequences.
[Order cDNA clone](#)

Click gene name

BCL2-associated X protein (BAX)

SELECTED PROTEIN SIMILARITIES

Comparison of sequences in UniGene with selected protein reference sequences. The alignments can suggest function of a gene.

	Reference Protein	Species	Id(%)	Len(a)
NP_004315.1	BCL2-associated X protein isoform beta	<i>H. sapiens</i>	100.0	
XP_001155879.1	PREDICTED: similar to Bax alpha	<i>P. troglodytes</i>	100.0	214
NP_031553.1	Bcl2-associated X protein	<i>M. musculus</i>	96.3	191
NP_001079104.1	BCL2-associated X protein	<i>X. laevis</i>	70.8	209
NP_571637.1	bcl2-associated X protein, a	<i>D. rerio</i>	59.8	189

Protein similarity info for BAX

GENE EXPRESSION

Tissues and development stages from this gene's sequences survey gene expression. Links to other NCBI expression resources.

[EST Profile:](#) Approximate expression patterns inferred from EST sources.
[\[Show more entries with profiles like this\]](#)

[GEO profiles:](#) Experimental gene expression data (Gene Expression Omnibus).

cDNA Sources: [pharynx](#); [embryonic tissue](#); [lung](#); [uterus](#); [lymph](#); [thymus](#); [mouth](#); [placenta](#); [pancreas](#); [testis](#); [ovary](#); [adipose](#); [blood](#); [adrenal gland](#); [ascites](#); [bladder](#); [bone](#); [brain](#); [colon](#); [esophagus](#); [heart](#); [intestine](#); [kidney](#); [liver](#); [muscle](#); [pancreas](#); [spleen](#); [stomach](#); [testis](#); [thymus](#); [trachea](#); [urinary bladder](#); [uterus](#); [vagina](#); [white blood cells](#); [yolk](#).

Gene expression information. Click on 'Expression Profile'

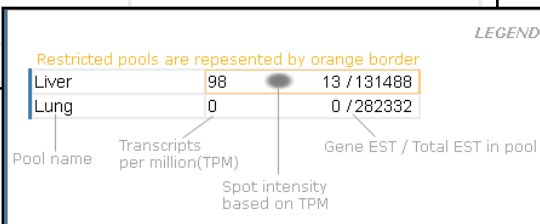
Breakdown by Body Sites

		Hs.624291
adipose tissue	0	0 / 13156
adrenal gland	29	1 / 33340
ascites	0	0 / 40056
bladder	0	0 / 30125
blood	16	2 / 124108

Tissue expression patterns for the human BAX gene.

Breakdown by Developmental Stage

		Hs.624291
embryoid body	70	5 / 70766
blastocyst	0	0 / 62333
fetus		
neonate		



B. WE2

EMBL-EBI

ArrayExpress

Services Research Training Industry About us

Home Experiments Arrays Submit Help About ArrayExpress

Feedback Login

ArrayExpress - functional genomics data

ArrayExpress is a database of functional genomics experiments that can be queried and the data downloaded. It includes gene expression data from microarray and high throughput sequencing studies. Data is collected to [MIAME](#) and [MINSEQE](#) standards. Experiments are submitted directly to ArrayExpress or are imported from the NCBI GEO database.

Data Content

Updated today at 06:00

- 43123 experiments
- 1223241 assays
- 17.77 TB of data

Search

Examples: E-MEXP-31, cancer, p53, Geuvadis

Advanced

Search Array
express for
“p53 RNA
-seq”

EMBL-EBI

ArrayExpress

Services Research Training Industry About us

Home Experiments Arrays Submit Help About ArrayExpress

Feedback Login

ArrayExpress > Search results for "p53 rna-seq"

ArrayExpress results for *p53 rna-seq*

Filter experiments

By organism: All organisms | By array: All arrays | By experiment type: All assays by molecule | All technologies

☐ ArrayExpress data only

Filter

16 experiments

Accession	Title	Type	Organism	Assays	Released	Processed
E-GEOD-45284	Regulation of constitutive and alternative splicing by PRMT5 reveals a role for Mdm4 pre-mRNA in sensing defects in the spliceosomal machinery (RNA-Seq)	RNA-seq of coding RNA	Mus musculus	12	13/09/2013	-
E-GEOD-31616	microRNA expression of cancer and matched adjacent tissues from 7 testicular germ cell tumors and 10 transitional cell carcinomas of bladder	RNA-seq of non coding RNA	Homo sapiens	34	01/09/2013	↓
E-GEOD-31614	mRNA expression of cancer and matched adjacent tissues of 7 testicular germ cell tumors and 10 transitional cell carcinomas of bladder	RNA-seq of coding RNA	Homo sapiens	34	01/09/2013	↓
E-MTAB-1683	RNA-seq of coding RNA from enterocytes taken from CK1 alpha deletion mutants (CK1 alpha KO), CK1 alpha/p53 double deletion mutants (CK1 alpha/p53 DKO) and CK1 alpha heterozygous mice (CK1 alpha Het) to elucidate the effect of senescence in mouse colorectal tumor	RNA-seq of coding RNA	Mus musculus	6	12/06/2013	-
E-GEOD-45785	Transcriptome-profiling (RNA-seq) and Ribosome-profiling (Ribo-seq) of BJ cells treated with Nutlin-3a, an MDM2 inhibitor, which induces p53.	other, RNA-seq of coding RNA	Homo sapiens	10	01/06/2013	↓

Select
Transcriptome and
Ribosome
profiling
(E-GEOD-45785)

EMBL-EBI

ArrayExpress

Services Research Training Industry About us

Home Experiments Arrays Submit Help About ArrayExpress

Feedback Login

ArrayExpress > Experiments > E-GEOD-45785

E-GEOD-45785 - Transcriptome-profiling (RNA-seq) and Ribosome-profiling (Ribo-seq) of BJ cells treated with Nutlin-3a, an MDM2 inhibitor, which induces p53.

Status: Released on 1 June 2013, last updated on 21 June 2013

Organism: Homo sapiens

Samples (10): [Click for detailed sample information and links to data](#)

Protocols (4): [Click for detailed protocol information](#)

Description: We applied in parallel RNA-Seq and Ribosome-profiling analyses to immortalized human primary BJ fibroblast cells in which p53 was induced by Nutlin-3a. RNA-seq, using Illumina HiSeq 2000, was applied to BJ cells treated with Nutlin-3a, at 5 timepoints: 0, 2, 4, 6, 19 hrs. Ribosome profiling was performed with Nutlin-3a, at 5 timepoints: 0, 2, 4, 6, 19 hrs.

Experiment types: other, RNA-seq of coding RNA

Contacts: [Ran Elkon <geo@ncbi.nlm.nih.gov>](#), Fabricio Loayza-Puch, Rani Elkon, Reuven Agami

Citation: [p53 induces transcriptional and translational programs to suppress cell proliferation and growth](#). Loayza-Puch F, Drost J, Rooijers K, Lopes R, Elkon R. [Europe PMC 23594524](#)

MINSEQE: ★ ★ ★ ★ ★

Exp. design: Protocols: Factors: Processed: Seq. reads:

Files: Investigation description: [E-GEOD-45785.idf.txt](#)
Sample and data relationship: [E-GEOD-45785.sdrf.txt](#)
Processed data (3): [E-GEOD-45785.processed.1.zip](#), [E-GEOD-45785.processed.2.zip](#), [E-GEOD-45785.processed.3.zip](#)
[Click to browse all available files](#)

Links: [GEO - GSE45785](#), [ENA - SRP020544](#)
[Send E-GEOD-45785 data to GENOME SPACE](#)

Information
on the study
including
files to
download.

GEO DataSets

Advanced Search

The information on this web site remains accessible; but, due to the lapse in government funding, the information may not be up to date, and the agency may not be able to respond to inquiries until appropriations are enacted. For updates regarding government operating status see USA.gov.

GEO DataSets

This database stores curated gene expression DataSets, as well as original Series and Platform records in the Expression Omnibus (GEO) repository. Enter search terms to locate experiments of interest. DataSet records additional resources including cluster tools and differential expression queries.

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

Send to:

Results: 14

- ☐ [Regulation of constitutive and alternative splicing by PRMT5 reveals a role for Mdm4 pre-mRNA in sensing defects in the spliceosomal machinery \(RNA-Seq\)](#)
 (Submitter supplied) Protein Arginine Methyltransferase 5 (PRMT5) is known to mediate epigenetic control on chromatin and to functionally regulate components of the splicing machinery. In this study we show that selective deletion of PRMT5 in different organs leads to cell cycle arrest and apoptosis. At the molecular level, PRMT5 depletion results in reduced methylation of Sm proteins, aberrant constitutive splicing and in the Alternative Splicing (AS) of specific mRNAs. [more...](#)
 Organism: Mus musculus
 Type: Expression profiling by high throughput sequencing
 Platform: GPL13112 12 Samples
 Download data: [GEO \(TXT\)](#), [SRA SRP019782](#)
 Series Accession: GSE45284 ID: 200045284
[PubMed](#) [Full text in PMC](#) [Similar studies](#)
- ☐ [Transcriptome-profiling \(RNA-seq\) and Ribosome-profiling \(Ribo-seq\) of BJ cells treated with Nutlin-3a, an MDM2 inhibitor, which induces p53.](#)

NCBI > GEO > **Accession Display**

Not logged in | [Login](#)

Scope: Self Format: HTML Amount: Quick GEO accession: GSE45785 [Go](#)

Series GSE45785 [Query DataSets for GSE45785](#)

Status	Public on Jun 01, 2013
Title	Transcriptome-profiling (RNA-seq) and Ribosome-profiling (Ribo-seq) of BJ cells treated with Nutlin-3a, an MDM2 inhibitor, which induces p53.
Organism	Homo sapiens
Experiment type	Expression profiling by high throughput sequencing Other
Summary	We applied in parallel RNA-Seq and Ribosome-profiling analyses to immortalized human primary BJ fibroblast cells in which p53 was induced by Nutlin-3a
Overall design	RNA-seq, using Illumina HiSeq 2000, was applied to BJ cells treated with Nutlin-3a, at 5 timepoints: 0, 2, 4, 6, 19 hrs Ribosome profiling was applied to BJ cells treated with Nutlin-3a, at 5 timepoints: 0, 2, 4, 6, 19 hrs
Contributor(s)	Loayza-Puch F , Elkon R , Agami R
Citation(s)	Loayza-Puch F, Drost J, Rooijers K, Lopes R et al. p53 induces transcriptional and translational programs to suppress cell proliferation and growth. <i>Genome Biol</i> 2013 Apr 17;14(4):R32. PMID: 23594524
Submission date	Apr 04, 2013
Last update date	Sep 30, 2013
Contact name	Ran Elkon
Organization name	The Netherlands Cancer Institute
Department	Gene Regulation
Lab	Reuven Agami
Street address	
City	Amsterdam
ZIP/Postal code	1066 CX
Country	Netherlands

Search GEO
for “p53
RNA-seq”

Select
Transcriptome
and
Ribosome
profiling
GSE45785

Similar
information
on the study
including
files to
download.

European Nucleotide Archive

The European Nucleotide Archive (ENA) provides a comprehensive record of the world's nucleotide sequencing information, covering raw sequencing data, sequence assembly information and functional annotation ... [more](#)

Access to ENA data is provided through the browser, through search tools, large scale file download and through the API.

Text search

p53 rna-seq

Search

[Advanced Search](#)

Search ENA
for "p53
RNA-seq"

Sequence Search

Search results for *p53 rna-seq*

Sequence

Sequence (Release) (3)

Transcriptome assembly contig (Release) (17)

Coding

Coding (Release) (20)

Read

Experiment (7)

Run (6)

Study

Study (17)

Study (Sequence) (21)

Study (17 results found)

View: XML

Download: 1 - 17 of 17 results in XML

Showing results 1 - 10 of 17 results

SRP006702 RNA Capture-Seq resolves the deep complexity of the human transcriptome (454)

SRP017378 GSE42509: Transcriptome-profiling (RNA-seq) in proliferation, quiescence, senescence... [more](#)

SRP019782 GSE45284: Splicing regulator PRMT5 links Mdm4 Alternative Splicing to the p53... [more](#)

SRP020544 GSE45785: Transcriptome-profiling (RNA-seq) and Ribosome-profiling (Ribo-seq) of BJ... [more](#)

SRP021219 GSE46240: Global genomic profiling of p53-regulated genes

SRP022870 GSE47041: Discovery of the p53 targetome in MCF7 cells from ChIP-Seq data

SRP022871 GSE47042: Discovery of the p53 targetome in MCF7 cells from RNA-seq data

SRP028963 GSE50064: p53 shapes genome-wide changes in small non-coding RNA expression during...

Select 'study'
and then

Transcriptome
and
Ribosome
profiling
SRP020544

Text search Advanced search Sequence search

Enter or paste text or ENA accession number:

Search

Upload file of accessions:

Browse... No file selected.

Search

Study: SRP020544 : GSE45785: Transcriptome-profiling (RNA-seq) and Ribosome-profiling (Ribo-seq) of BJ cells treated with [more](#)

View: XML

Download: XML

[Send Feedback](#)

Submitting Centre

GEO

Study Type

Other

Read Count

Base Count

Broker Name

NCBI

Abstract

Summary: We applied in parallel RNA-Seq and Ribosome-profiling analyses to immortalized human primary BJ fibroblast cells in which p53 was induced by Nutlin-3a. Overall Design: RNA-seq, using Illumina HiSeq 2000, was applied to BJ cells treated with Nutlin-3a, at 5 timepoints: 0, 2, 4, 6, 19 hrs. Ribosome profiling was applied to BJ cells treated with Nutlin-3a, at 5 timepoints: 0, 2, 4, 6, 19 hrs.

Navigation Read Files Attributes

Download files

View: TEXT

Download: TEXT

Select columns

Showing results 1 - 10 of 10 results

Study accession	Secondary study accession	Sample accessions	Experiment accession	Run accession	Scientific name	Instrument model	Library layout	Fastq files (ftp)	Fastq files (galaxy)	Submitted files (ftp)	Submitted files (galaxy)	CoL taxonomy
SRP020544	SRP020544	SRS408262	SRX260240	SRR810095	Homo sapiens	Illumina HiSeq 2000	PAIRED	File 1 File 2	File 1 File 2	Not available	Not available	
SRP020544	SRP020544	SRS408263	SRX260241	SRR810096	Homo sapiens	Illumina HiSeq 2000	PAIRED	File 1 File 2	File 1 File 2	Not available	Not available	
SRP020544	SRP020544	SRS408264	SRX260242	SRR810097	Homo sapiens	Illumina HiSeq 2000	PAIRED	File 1 File 2	File 1 File 2	Not available	Not available	
SRP020544	SRP020544	SRS408265	SRX260243	SRR810098	Homo sapiens	Illumina HiSeq 2000	PAIRED	File 1 File 2	File 1 File 2	Not available	Not available	
SRP020544	SRP020544	SRS408266	SRX260244	SRR810099	Homo sapiens	Illumina HiSeq 2000	PAIRED	File 1 File 2	File 1 File 2	Not available	Not available	
SRP020544	SRP020544	SRS408267	SRX260245	SRR810100	Homo sapiens	Illumina HiSeq 2000	SINGLE	File 1	File 1	Not available	Not available	

To export files
to Galaxy
click the
appropriate
files.

Click 'Geo profiles' from Entrez Gene page

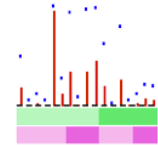
Results: 1 to 20 of 3033

<< First < Prev Page 1 of 152 Next > Last >>

☐ [BAX - Type 2 diabetes and role of hepatokines](#)

1. Annotation: BAX, BCL2-associated X protein
Organism: Homo sapiens
Reporter: GPL570, 211833_s_at (ID_REF), GDS3883, 581 (Gene ID), U19599
DataSet type: Expression profiling by array, transformed count, 17 samples
ID: 71244650

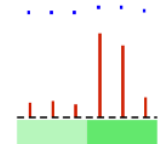
[GEO DataSets](#) [Gene](#) [UniGene](#) [Profile neighbors](#) [Chromosome neighbors](#) [Sequence neighbors](#) [Homologene neighbors](#)



☐ [BAX - Diabetic nephropathy](#)

2. Annotation: BAX, BCL2-associated X protein
Organism: Homo sapiens
Reporter: GPL8300, 1998_i_at (ID_REF), GDS961, 581 (Gene ID), U19599
DataSet type: Expression profiling by array, count, 6 samples
ID: 8551707

[GEO DataSets](#) [Gene](#) [UniGene](#) [Profile neighbors](#) [Chromosome neighbors](#) [Sequence neighbors](#) [Homologene neighbors](#)



Click GDS961 entry

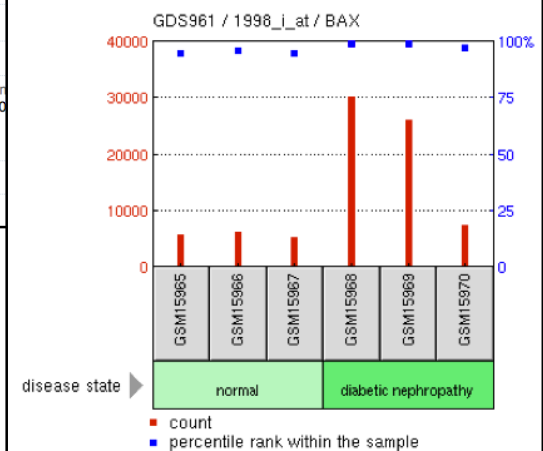
Click graphic

DataSet Record GDS961: Expression Profiles Data Analysis Tools			
Title:	Diabetic nephropathy		
Summary:	Comparison of glomeruli from kidneys with diabetic nephropathy (DN) and glomeruli of healthy individuals. Progression of DN may be due to diminished tissue repair capacity.		
Organism:	<i>Homo sapiens</i>		
Platform:	GPL8300: [HG_U95Av2] Affymetrix Human Genome U95 Version 2 Array		
Citation:	Baelde HJ, Eikmans M, Doran PP, Lappin DW et al. Gene expression profiling in glomeruli of human kidneys with diabetic nephropathy. <i>Am J Kidney Dis</i> 2004 Apr;43(4):636-50. PMID: 15042541		
Reference Series:	GSE1009	Sample count:	6
Value type:	count	Series published:	2004/02/02

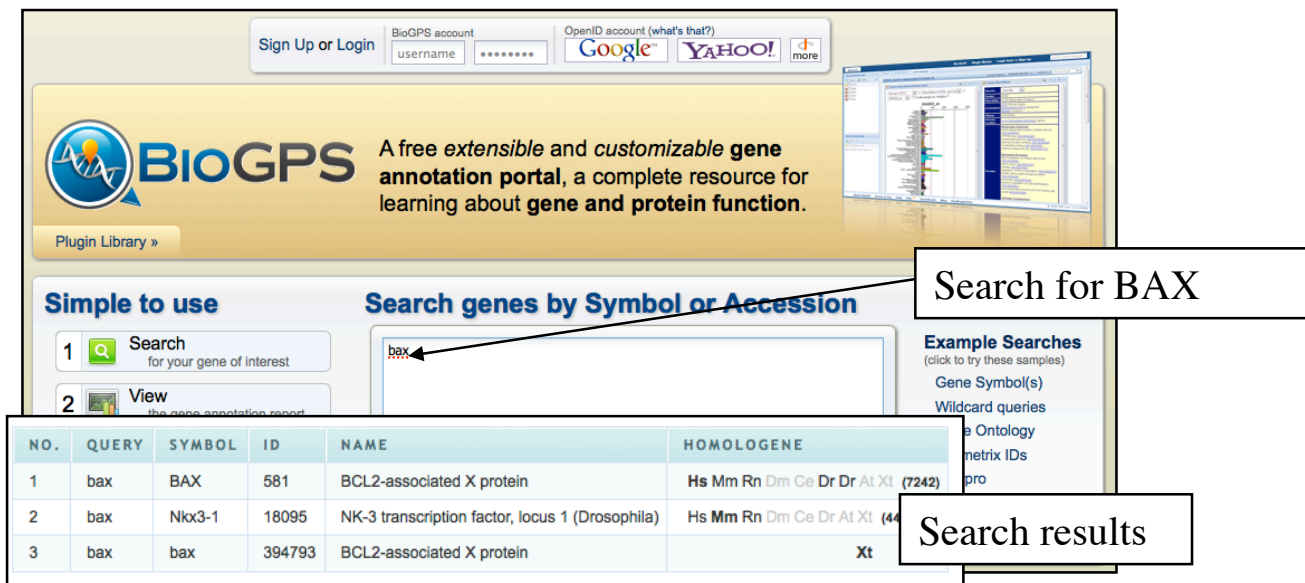
Profile GDS961 / 1998_i_at / BAX

Title Diabetic nephropathy

Organism Homo sapiens



BioGPS: <http://biogps.gnf.org>

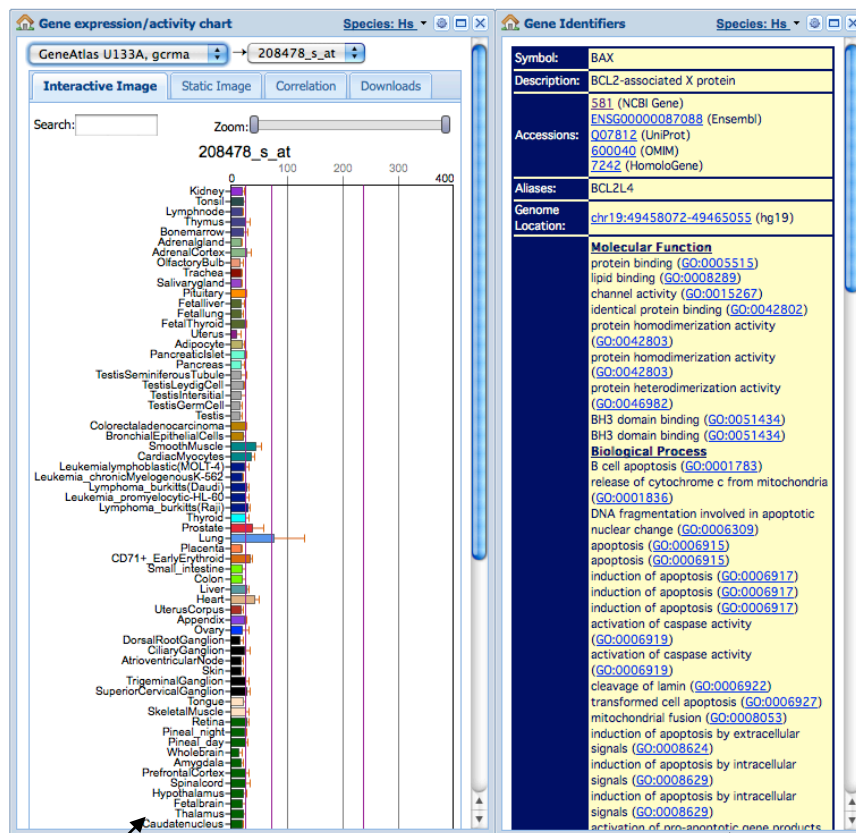


The BioGPS homepage features a search bar with the text "Search for your gene of interest" and a "View the gene annotation report" button. Below the search bar is a table of search results. An arrow points from the text "Search for BAX" to the search bar, and another arrow points from the text "Search results" to the table.

Search for BAX

Search results

NO.	QUERY	SYMBOL	ID	NAME	HOMOLOGENE
1	bax	BAX	581	BCL2-associated X protein	Hs Mm Rn Dm Ce Dr At Xt (7242)
2	bax	Nkx3-1	18095	NK-3 transcription factor, locus 1 (Drosophila)	Hs Mm Rn Dm Ce Dr At Xt (44)
3	bax	bax	394793	BCL2-associated X protein	Xt



Microarray expression information

Gene info

View BAX protein structure info in PDB
<http://www.rcsb.org/pdb/home/home.do>

RCSB PDB PROTEIN DATA BANK **PDB-101** A MEMBER OF THE PDB
An Information Portal to Biological Macromolecular Structures
As of **Tuesday Nov 15, 2011 at 4 PM PST** there are **77243** Structures | [PDB Statistics](#) | [RSS](#) | [Facebook](#) | [Twitter](#)

Search | All Categories: ☐ All Categories ☐ Author ☐ Macromolecule ☒ Sequence ☐ Ligand ☐ ?

Search for human BAX in PDB.
Safest way is to do a sequence search.
Paste sequence from human BAX
protein

85 Structure Entity Hits **41 Citations** **38 Ligand Hits**

Structure Entity Query: Your search is an entity-based query. The 85 entities (4)

Query Parameters:
Sequence Search:
MDGSGEQPRGGGPTSSEQIMKTGALLQGFQDRAGRMGGEAPELALDPV PQDASTKKLSECLKRIGDELDSNMELQRMIAAVDTDSPREVFFRVAADM
SDGNFNWGRVVALFYFASKVLKALCTKVPPELIRTINGWTLDFLRERLLG WIQDQGGWGLLSYFGTPTWQTVTIFVAGVLTASLTINWKKMG
Expectation Value = 0.01, Search Tool = blast, Mask Low Complexity=yes

Query Refinements

Organism
• Homo sapiens (62)
• Mus musculus (21)
• Rattus norvegicus (3)
• Schistosoma japonicum (1)
• Escherichia coli (1)

Taxonomy
• Eukaryota (84)
• Bacteria (1)

Experimental Method
• X-RAY (59)
• Solution NMR (26)

X-Ray Resolution
• less than 1.5 Å (2)
• 1.5 - 2.0 Å (14)
• 2.0 - 2.5 Å (29)
• 2.5 - 3.0 Å (13)

Release Date
• before 2000 (4)
• 2000 - 2005 (14)
• 2005 - 2010 (42)
• 2010 - today (25)
• this year (5)
• more choices...

Polymer Type
• Protein (85)

Sequence length
• less than 50 (3)
• 100 - 200 (70)
• 200 - 300 (12)
• more choices...

SCOP Classification
• Membrane and cell surface prote ... (33)

☒ Display/Download:

Summary of results

Top hit : Structure 1F16

Displaying results 1 - 25 of 85 total | Page 1 of 4 | Jump to page:

☒ **1F16** **SOLUTION STRUCTURE OF A PRO-APOPTOTIC PROTEIN BAX**

Chain(s): A

Authors: Suzuki, M., Youle, R.J., Tjandra, N.

Release Date: 2000-11-22 **Classification:** Apoptosis

Experiment: SOLUTION NMR

Compound: 1 Polymer [[Display Full Polymer Details](#) | [Display for All Results](#)]

Alignment: E-Value: 5.27988E-109 Score: 1001 [[Display Full Alignment](#) | [Display for All Results](#)]

Citation: Structure of Bax: coregulation of dimer formation and intracellular localization. (2000) Cell(Cambridge,Mass.) 103: 645-654 [[Display Full Abstract](#) | [Display for All Results](#)]

SOLUTION STRUCTURE OF A PRO-APOPTOTIC PROTEIN BAX

DOI:10.2210/pdb1f16/pdb

Primary Citation

Structure of Bax: coregulation of dimer formation and intracellular localization.

Suzuki, M., Youle, R.J., Tjandra, N.

Journal: (2000) Cell(Cambridge,Mass.) **103**: 645-654

PubMed: [11106734](#)

[Search Related Articles in PubMed](#)

PubMed Abstract:

Apoptosis is stimulated by the insertion of Bax from the cytosol into mitochondrial membranes. The solution structure of Bax, including the putative transmembrane domain at the C terminus, was determined in order to understand the regulation of its subcellular location.... [\[Read More & Search PubMed Abstracts \]](#)

Molecular Description

Classification: Apoptosis

Structure Weight: 21204.50

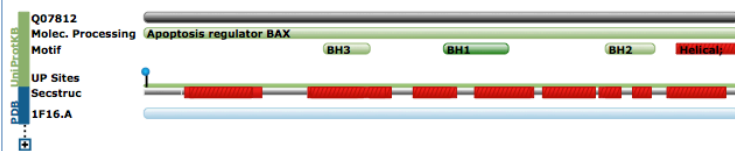
Molecule: PROTEIN (APOPTOSIS REGULATOR BAX, MEMBRANE ISOFORM ALPHA)

Polymer: 1 Type: protein Length: 192

Chains: A

Organism: *Homo sapiens*

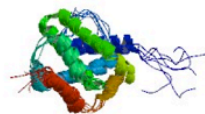
UniProtKB: [Protein Feature View](#) | [Search PDB](#) | [Q07812](#)



1F16

[Display Files](#)
[Download Files](#)
[Share this Page](#)

Structure Image



[View in 3D](#)

[More Images...](#)

Downloadable viewers:

[Simple Viewer](#) [Protein Workshop](#)
[Ligand Explorer](#) [Kiosk Viewer](#)

MyPDB Personal Annotations

To save personal annotations, please [login](#) to your MyPDB account.

Deposition Summary

Authors: Suzuki, M., Youle, R.J., Tjandra, N.

Deposition: 2000-05-18
 Release: 2000-11-22

[View in Jmol](#)



Jmol S

The Genome Access Course

UCSC Genome Browser



The UCSC Genome Browser

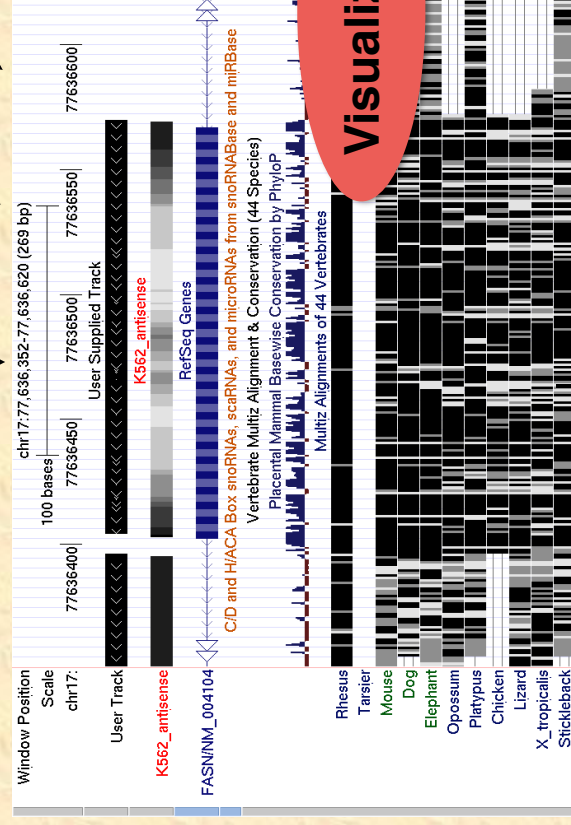
Assaf Gordon
Erich Lab
New York Genome Center

The Genome Access Course 2015

What is a "Genome Browser" ?



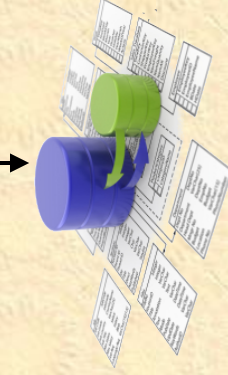
Aggregation



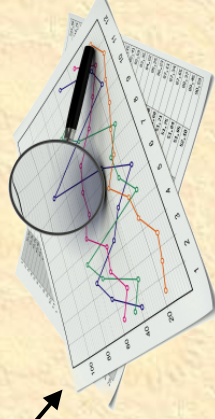
Visualization



Analysis

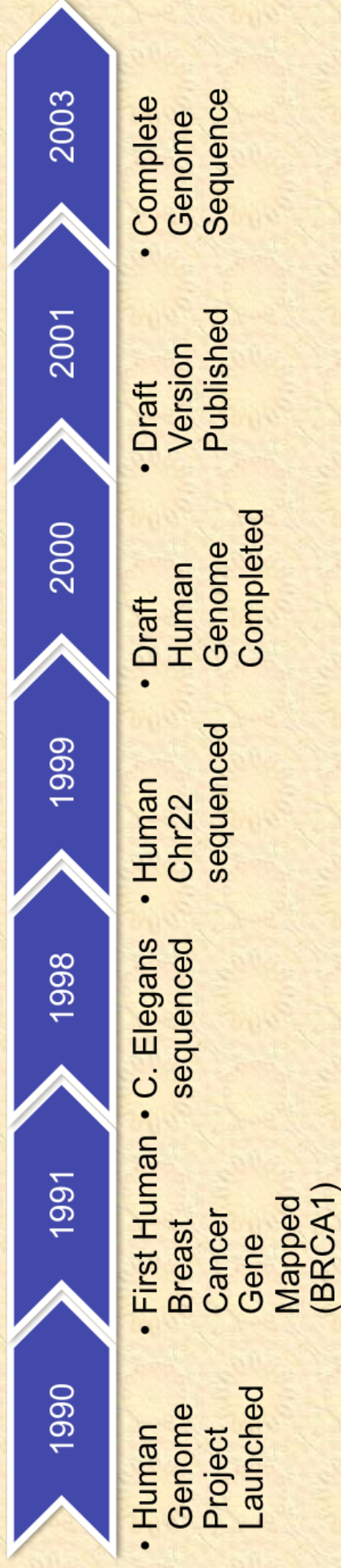


Data Export



Exploration

So Much Data



• The Reference Human Genome has ~3.3 billion bases

- 5 'versions' of the reference genome since 2003

- million of features (e.g genes, isoforms, miRNAs) from multiple sources

- >90 organisms, 242 reference genomes

- >52 TB to store all the data (1TB=1,000GB=1,000,000MB).

- Periodically updated

A Taste of the UCSC Genome Browser

Navigation

Tracks

Sequence
Search
(blat)

Sessions,
Data Sharing

Table
Browser

Custom
Tracks

A Taste of the UCSC Genome Browser

Navigation

Tracks

Sequence
Search
(blat)

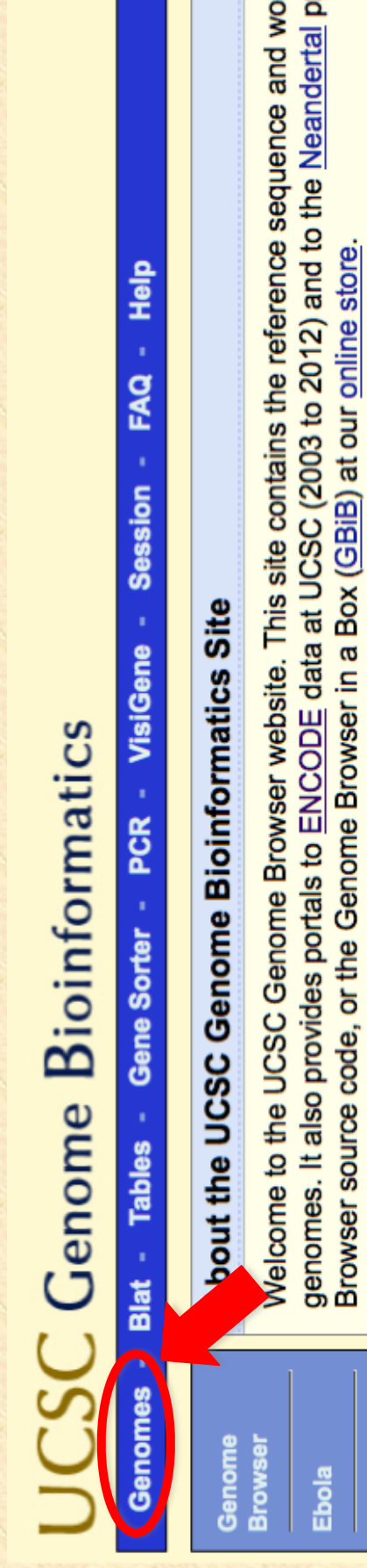
Sessions,
Data Sharing

Table
Browser

Custom
Tracks

Navigation

- Visit <http://genome.ucsc.edu>



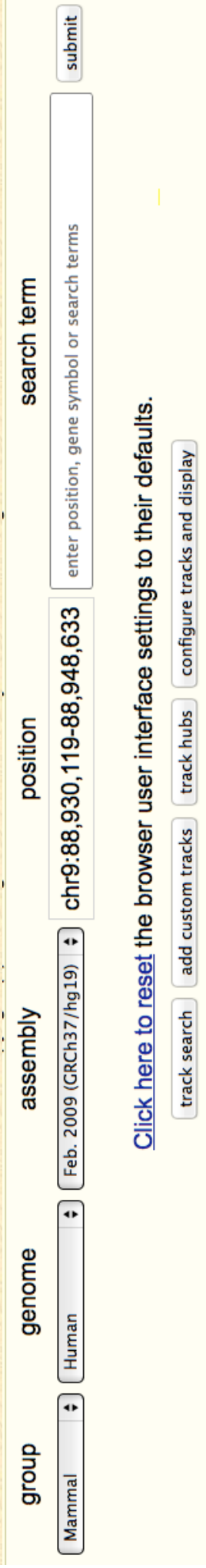
UCSC Genome Bioinformatics

[Genomes](#) - [Blat](#) - [Tables](#) - [Gene Sorter](#) - [PCR](#) - [VisiGene](#) - [Session](#) - [FAQ](#) - [Help](#)

About the UCSC Genome Bioinformatics Site

Welcome to the UCSC Genome Browser website. This site contains the reference sequence and whole genomes. It also provides portals to [ENCODE](#) data at UCSC (2003 to 2012) and to the [Neanderthal](#) project. You can view the genome source code, or the Genome Browser in a Box ([GBiB](#)) at our [online store](#).

- Select Organism, Assembly, Locus or Gene



group: Mammal | genome: Human | assembly: Feb. 2009 (GRCh37/hg19) | position: chr9:88,930,119-88,948,633 | search term: enter position, gene symbol or search terms

submit

[Click here to reset the browser user interface settings to their defaults.](#)

track search | add custom tracks | track hubs | configure tracks and display

Gene Search Results

UCSC Genes

[BRCA1 \(uc002ict.3\)](#) at [chr17:41196312-41277500](#) - Homo sapiens breast cancer 1, early onset (BRCA1), transcript variant 2, mRNA.
[BRCA1 \(uc010wht.1\)](#) at [chr17:41243452-41277500](#) - Homo sapiens breast cancer 1, early onset (BRCA1), transcript variant 1, mRNA.
[BRCA1 \(uc010whs.1\)](#) at [chr17:41243452-41277468](#) - Homo sapiens breast cancer 1, early onset (BRCA1), transcript variant 1, mRNA.
[BRCA1 \(uc010whr.1\)](#) at [chr17:41215350-41277468](#) - Homo sapiens breast cancer 1, early onset (BRCA1), transcript variant 3, mRNA.
[BRCA1 \(uc010whq.1\)](#) at [chr17:41215350-41256973](#) - Homo sapiens breast cancer 1, early onset (BRCA1), transcript variant 3, mRNA.
[BRCA1 \(uc010whp.2\)](#) at [chr17:41196312-41322420](#) - Homo sapiens breast cancer 1, early onset (BRCA1), transcript variant 4, mRNA.

RefSeq Genes

[BRCA1](#) at [chr17:41196312-41277340](#) - (NR_027676)
[BRCA1](#) at [chr17:41196312-41277500](#) - (NM_007294) breast cancer type 1 susceptibility protein isoform 1
[BRCA1](#) at [chr17:41196312-41277500](#) - (NM_007300) breast cancer type 1 susceptibility protein isoform 2
[BRCA1](#) at [chr17:41196312-41277468](#) - (NM_007297) breast cancer type 1 susceptibility protein isoform 3
[BRCA1](#) at [chr17:41196312-41276132](#) - (NM_007298) breast cancer type 1 susceptibility protein isoform 4
[BRCA1](#) at [chr17:41196312-41277468](#) - (NM_007299) breast cancer type 1 susceptibility protein isoform 5

Non-Human RefSeq Genes

[BRCA1](#) at [chr17:41197695-41276113](#) - (NM_001114949) breast cancer type 1 susceptibility protein homolog
[BRCA1](#) at [chr17:41197695-41276113](#) - (NM_001045493) breast cancer type 1 susceptibility protein homolog
[BRCA1](#) at [chr17:41197695-41276113](#) - (NM_001302103) breast cancer 1, early onset
[BRCA1](#) at [chr17:41197709-41276113](#) - (NM_001013416) breast cancer type 1 susceptibility protein homolog
[BRCA1](#) at [chr17:41197719-41276113](#) - (NM_178573) breast cancer type 1 susceptibility protein homolog
[BRCA1](#) at [chr17:41196320-41277431](#) - (NM_009764) breast cancer type 1 susceptibility protein homolog

Basic Gene Annotation Set from GENCODE Version 19

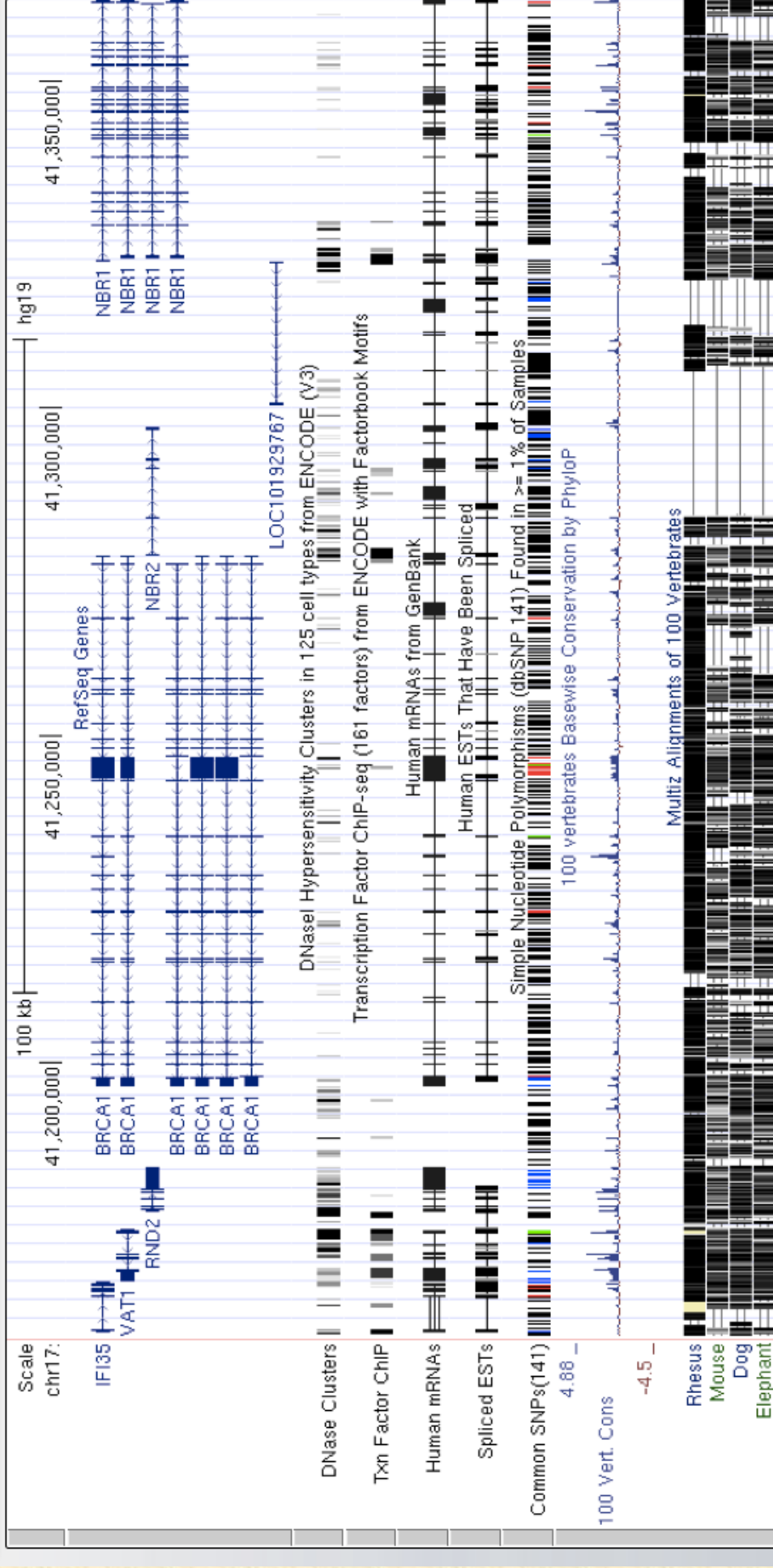
[BRCA1](#) at [chr17:41196312-41277387](#)
[BRCA1](#) at [chr17:41196313-41276132](#)
[BRCA1](#) at [chr17:41196313-41277468](#)
[BRCA1](#) at [chr17:41196313-41277468](#)
[BRCA1](#) at [chr17:41196313-41277500](#)
[BRCA1](#) at [chr17:41196313-41277500](#)

Navigation

UCSC Genome Browser on Human Feb. 2009 (GRCh37/hg19) Assembly

move <<< << < > >> >>> zoom in 1.5x 3x 10x base zoom out 1.5x 3x 10x 100x
chr17:41,158,329-41,362,622 204,294 bp. enter position, gene symbol or search terms go

chr17 (q21.31) 13.3 13.2 p13.1 17p12 17p11.2 17q12 17q11.2 17q22 23.2 24.2 q24.3 q25.1 17q25.3



A Taste of the UCSC Genome Browser

Navigation

Tracks

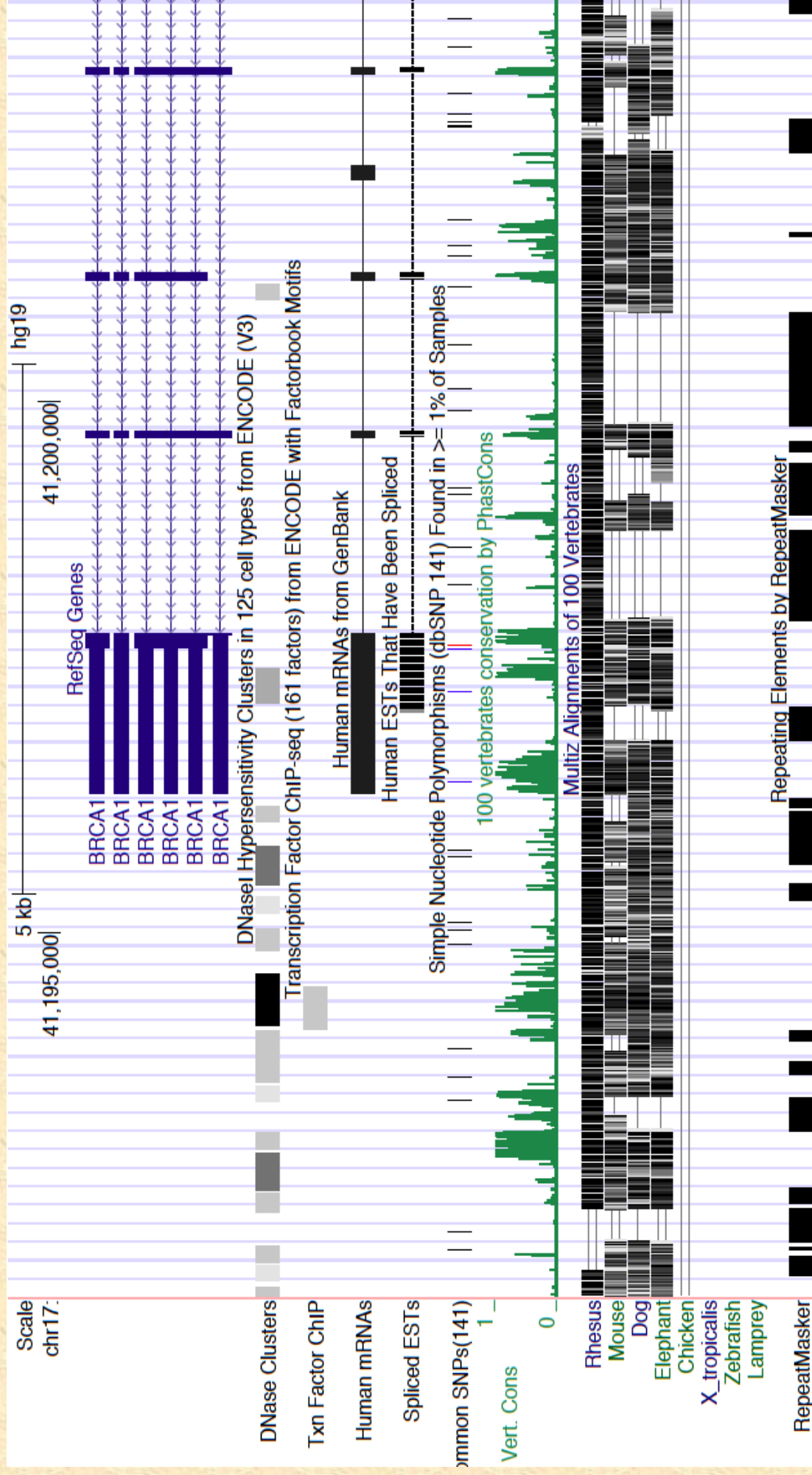
Sequence
Search
(blat)

Sessions,
Data Sharing

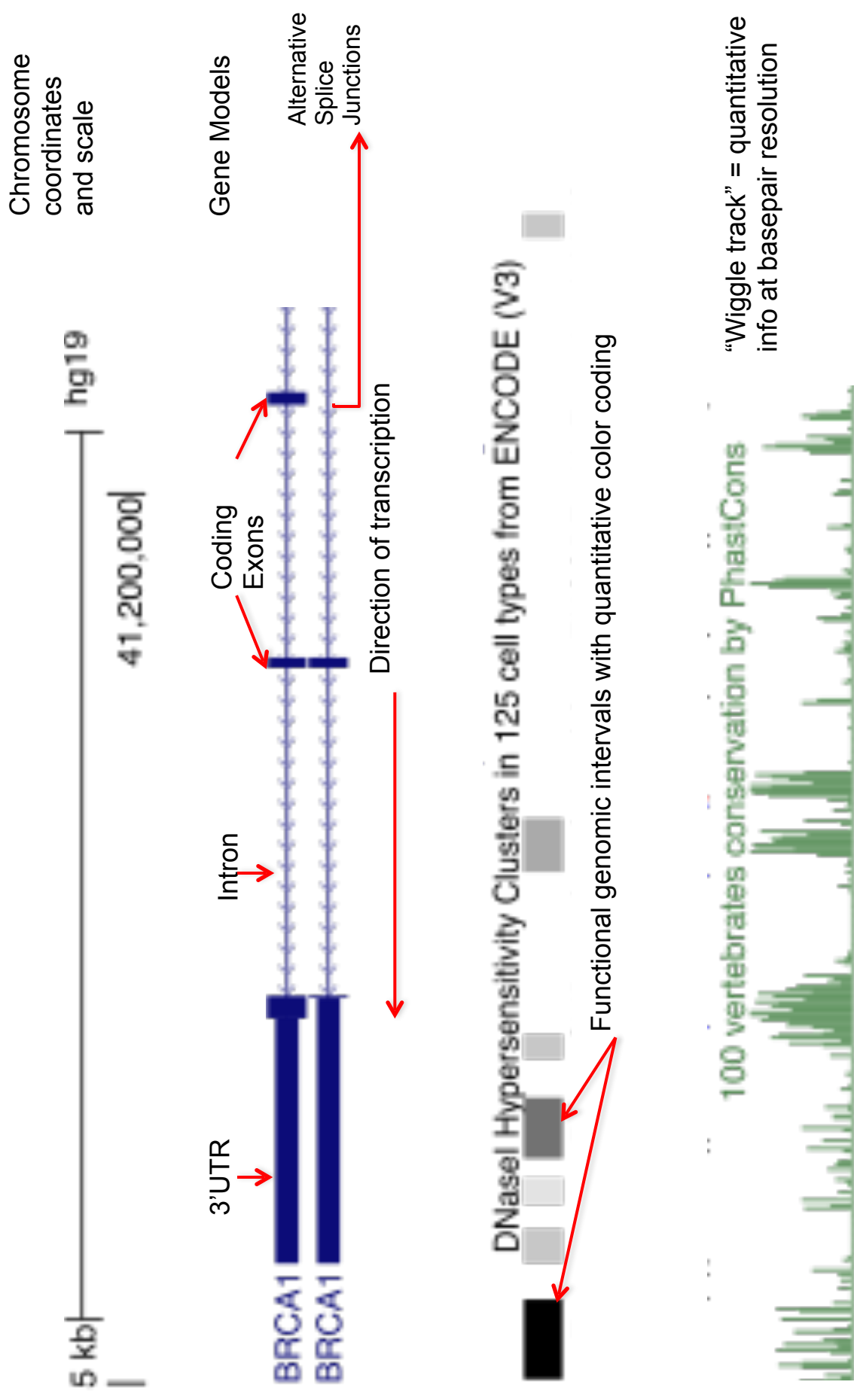
Table
Browser

Custom
Tracks

Tracks



Tracks - Display Conventions



Selecting Tracks

move start < 2.0 >

BRCA1 BRCA1

move end < 2.0 >

track search default tracks default order hide all add custom tracks track hubs configure reverse resize refresh

collapse all expand all refresh

Click on a feature for details. Click or drag in the base position track to zoom in. Click side bars for track options. Drag side bars or labels up or down to reorder tracks. Drag tracks left or right to new position.

Use drop-down controls below and press refresh to alter tracks displayed.
Tracks with lots of items will automatically be displayed in more compact modes.

Mapping and Sequencing

Base Position	Assembly	BAC End Pairs	BU ORChID	Chromosome Band	deCODE
dense	hide	hide	hide	hide	Recomb
ENCODE					hide
Pilot	18 FISH Clones	18 Fosmid End Pairs	GC Percent	GRC Incident	
[No data-chr17]	hide	hide	hide	hide	
GRC Map	GRC Patch Release	Hg18 Diff	Hg38 Diff	Hi Seq Depth	INSDC
Contigs	hide	hide	hide	hide	hide
LRG Regions	Map Contigs	Mappability	Recomb Rate	Restr Enzymes	Short Match
hide	hide	hide	hide	hide	hide
STS Markers	Wiki Track				
hide	hide				

Genes and Gene Predictions

refresh

Phenotype and Literature

refresh

mRNA and EST

refresh

Expression

refresh

Regulation

refresh

Comparative Genomics

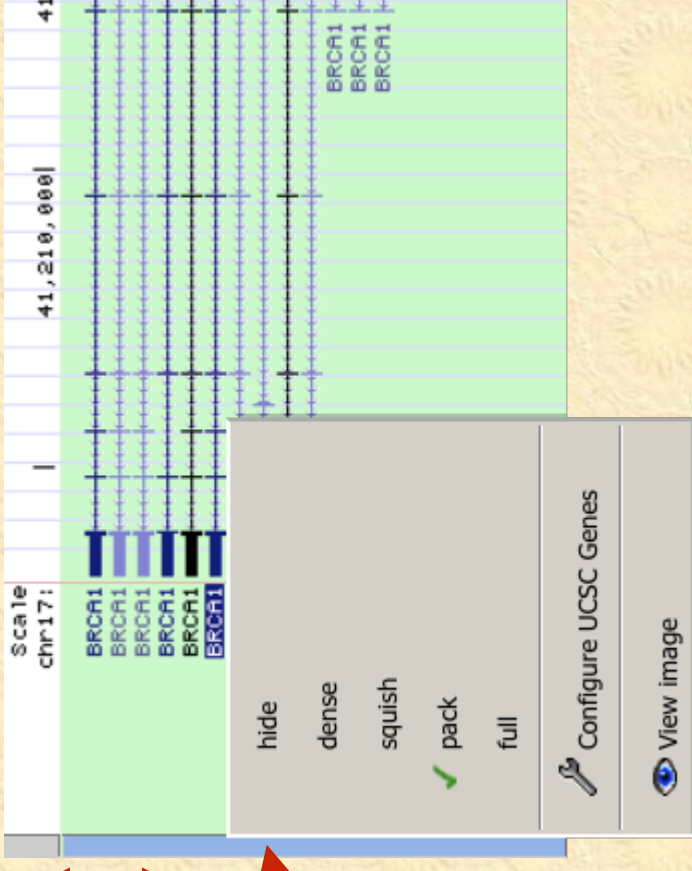
refresh

Tracks

Groups

Track Display Options

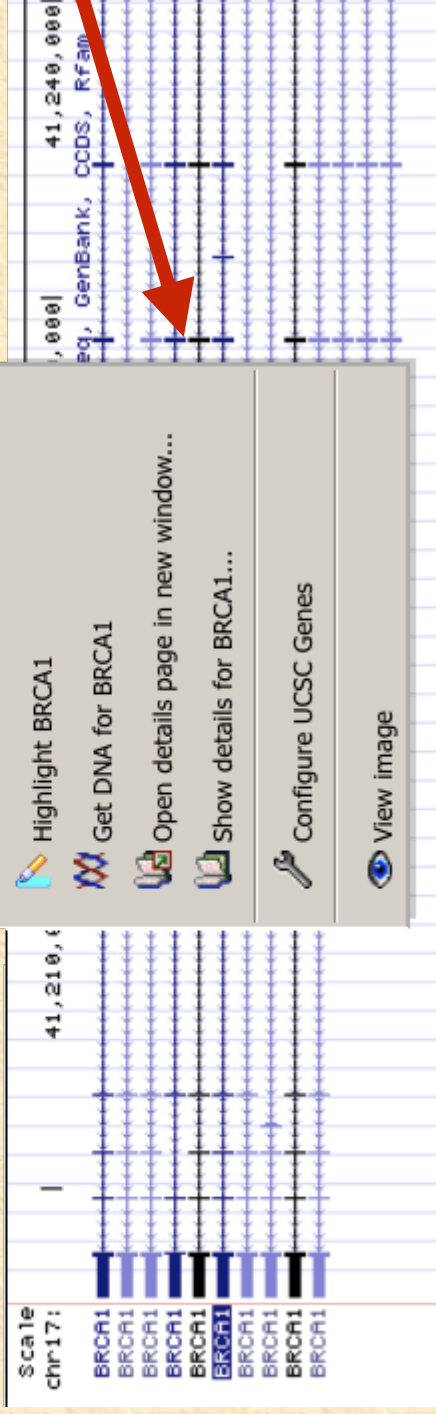
Drag gray/blue bars up or down to reorder tracks



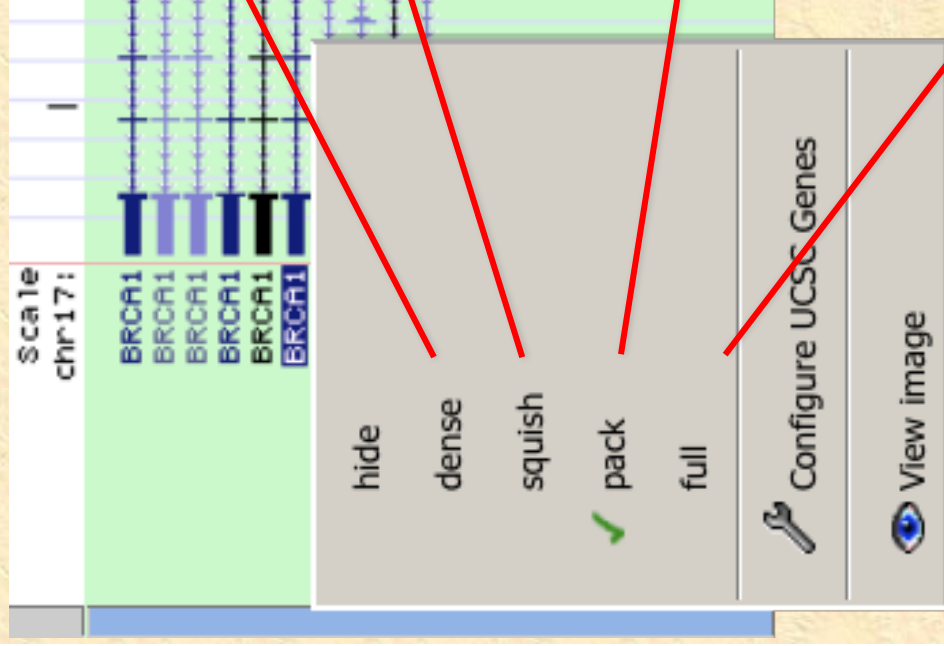
Right or Ctrl click on track to show display options.



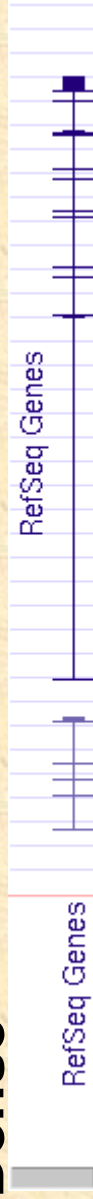
Right or Ctrl click on a gene to get more information.



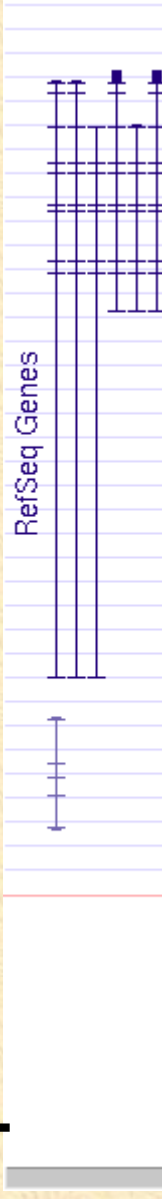
Track Display Options



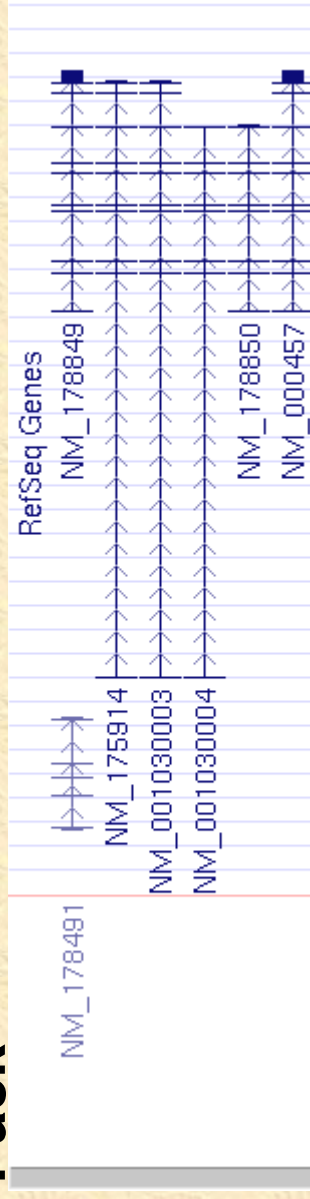
Dense



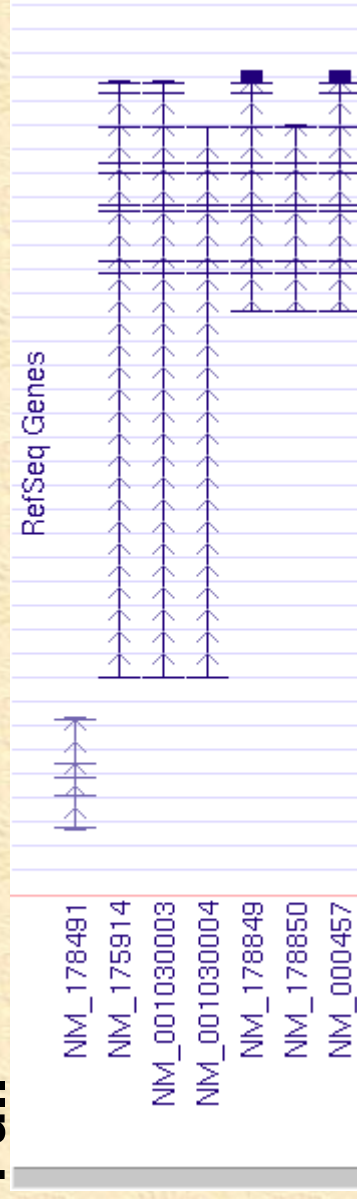
Squish



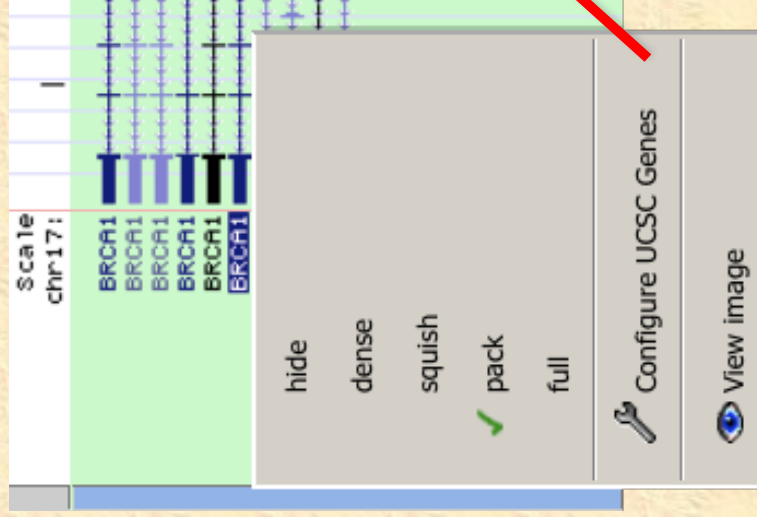
Pack



Full



Track Configuration



✕

UCSC Genes Track Settings

UCSC Genes (RefSeq, GenBank, CCDS, Rfam, tRNAs & Comparative Genomics)

Display mode: pack ▾

Label: ☒ gene symbol ☐ UCSC Known Gene ID ☐ UniProt Display ID ☐ OMIM ID

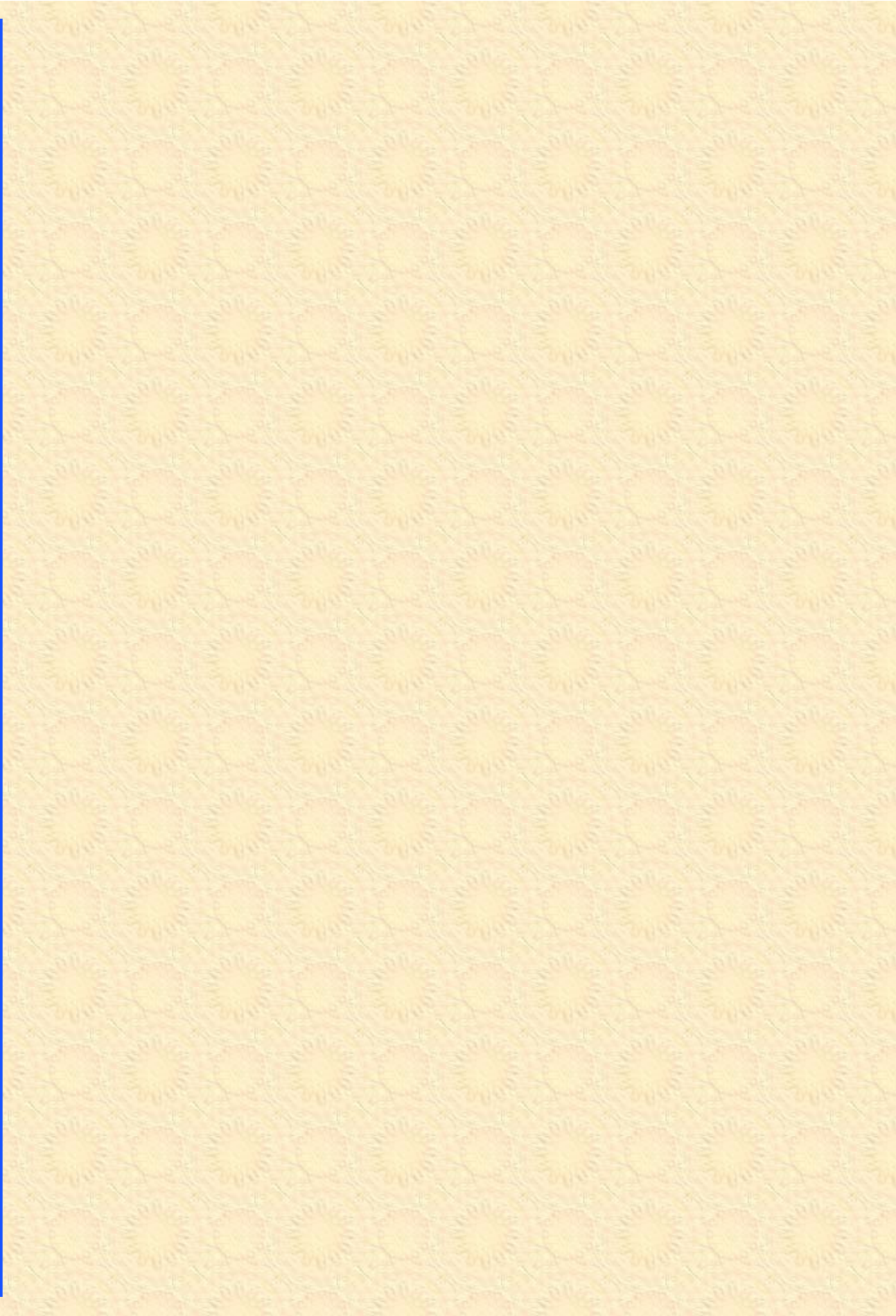
Show: ☒ non-coding genes ☒ splice variants

Color track by codons: genomic codons ▾ [Help on codon coloring](#)

Show codon numbering: ☐

OK

Demo



A Taste of the UCSC Genome Browser

Navigation

Tracks

Sequence
Search
(blat)

Sessions,
Data Sharing

Table
Browser

Custom
Tracks

Sequence Search - Blat

The screenshot shows the UCSC Genome Browser interface. The top navigation bar includes links for Genomes, Genome Browser, Tools, Mirrors, Downloads, My Data, Help, and About Us. The main content area is titled 'Human (Homo sapiens) Genome Browser'. On the left, a sidebar menu lists various tools: Blat, Table Browser, Variant Annotation Integrator, Gene Sorter, Genome Graphs, In-Silico PCR, LiftOver, VisiGene, and Other Utilities. A red arrow points to the 'Blat' link. Below the menu, there are input fields for 'group' (set to 'Mammal') and 'genome' (set to 'Human'). The main content area displays genomic data, including a position range of 'chr12:51,385,091-59,406,178' and a description of the browser user interface settings.

The February 2009 human reference sequence (GRCh37) was produced by the [Genome Reference Consortium](#). For information about this assembly, see [GRCh37](#) in the NCBI Assembly database.

Sequence Search - Blat

Human BLAT Search

BLAT Search Genome

Genome: Assembly: Query type: Sort output: Output type:

```
CCTCCACATCAACAACCTTAATGAGCTCCTTGAGATTTTAGTTTCTA
TTCTGAAGACTCCCAGAGCAACTGTGCAT
```

Paste in a query sequence to find its location in the genome. Multiple sequences may be searched if separated by lines starting with '>' followed by the sequence name.

File Upload: Rather than pasting a sequence, you can choose to upload a text file containing the sequence.

Upload sequence: No file selected.

One Sequence

Sequence Search - Blat

Human BLAT Search

BLAT Search Genome

Genome: Assembly: Query type: Sort output: Output type:

```
>Seq1
CCTCCACATCAACAACCTTAATGAGCTCCTTGAGATTTTAGTTTCTA
TTCTGAAGACTCCAGAGCAACTGTGCAT
>Seq2
AGTTGTCCAGTGTTGTGTTTCAGTGTACTAGTGTATCGAGTCGTTTAG
CGATTAAAGCCAGATCGATAACGCTAGCTACTAGCTA
>Seq3
TTGAGTGTGATGATTTTGAGCGCCCGATAGCTCGATCGAGCTAGCATCGAC
```

Paste in a query sequence to find its location in the the genome. Multiple sequences may be searched if separated by lines starting with '>' followed by the sequence name.

File Upload: Rather than pasting a sequence, you can choose to upload a text file containing the sequence.

Upload sequence: No file selected.

**Multiple
Sequences
(FASTA Format)**

Sequence Search - Blat

Human BLAT Search

BLAT Search Genome

Genome: Assembly: Query type: Sort output: Output type:

Paste in a query sequence to find its location in the the genome. Multiple sequences may be searched if separated by lines starting with '>' followed by the sequence name.

File Upload: Rather than pasting a sequence, you can choose to upload a text file containing the sequence.

Upload sequence: No file selected.

Select an
existing
file on your
computer



Sequence Search - Blat

Human BLAT Search

BLAT Search Genome

Genome: Assembly: Query type: Sort output: Output type:

```
CCTCCACATCAACAACCTTAATGAGCTCCTTGAGATTTTAGTTTCTA
TTCTGAAGACTCCCAGAGCAACTGTGCAT
```

Paste in a query sequence to find its location in the the genome. Multiple sequences may be searched if separated by lines starting with '>' followed by the sequence name.

File Upload: Rather than pasting a sequence, you can choose to upload a text file containing the sequence.

Upload sequence:

No file selected.

Sequence Search - Blat Results

Human BLAT Results

BLAT Search Results

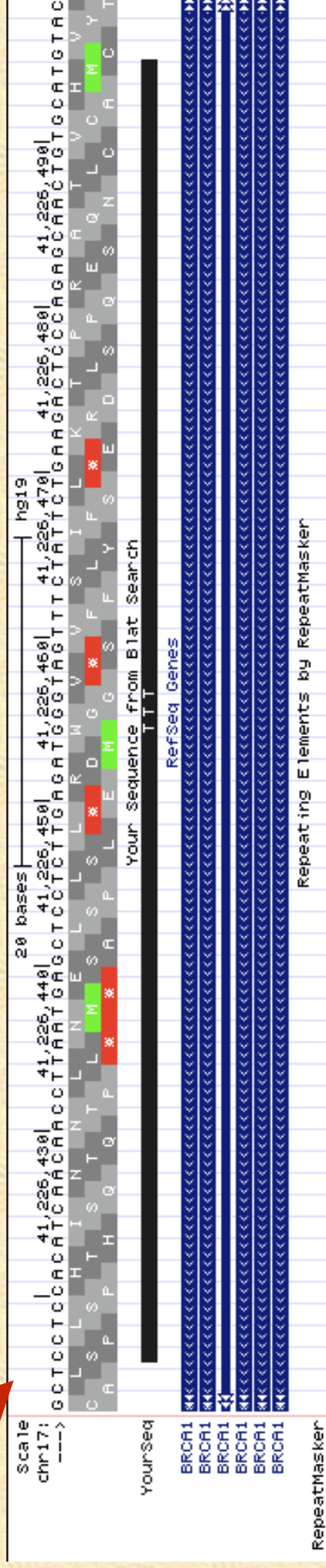
ACTIONS	QUERY	SCORE	START	END	QSIZE	IDENTITY	CHRO	STRAND	START	END	SPAN
browser details	YourSeq	73	1	79	79	96.3%	17	+	41226417	41226495	79
browser details	YourSeq	25	41	66	79	100.0%	5	+	116751191	116751353	163
browser details	YourSeq	20	30	49	79	100.0%	6	+	44372876	44372895	20
browser details	YourSeq	20	37	56	79	100.0%	11	+	46051786	46051805	20

Sequence Search - Blat Results

Human BLAT Results

BLAT Search Results

ACTIONS	QUERY	SCORE	START	END	QSIZE	IDENTITY	CHRO	STRAND	START	END	SPAN
browser details	YourSeq	73	1	79	79	96.3%	17	+	41226417	41226495	79
browser details	YourSeq	25	41	66	79	100.0%	5	+	116751191	116751353	163
browser details	YourSeq	20	30	49	79	100.0%	6	+	44372876	44372895	20
browser details	YourSeq	20	37	56	79	100.0%	11	+	46051786	46051805	20

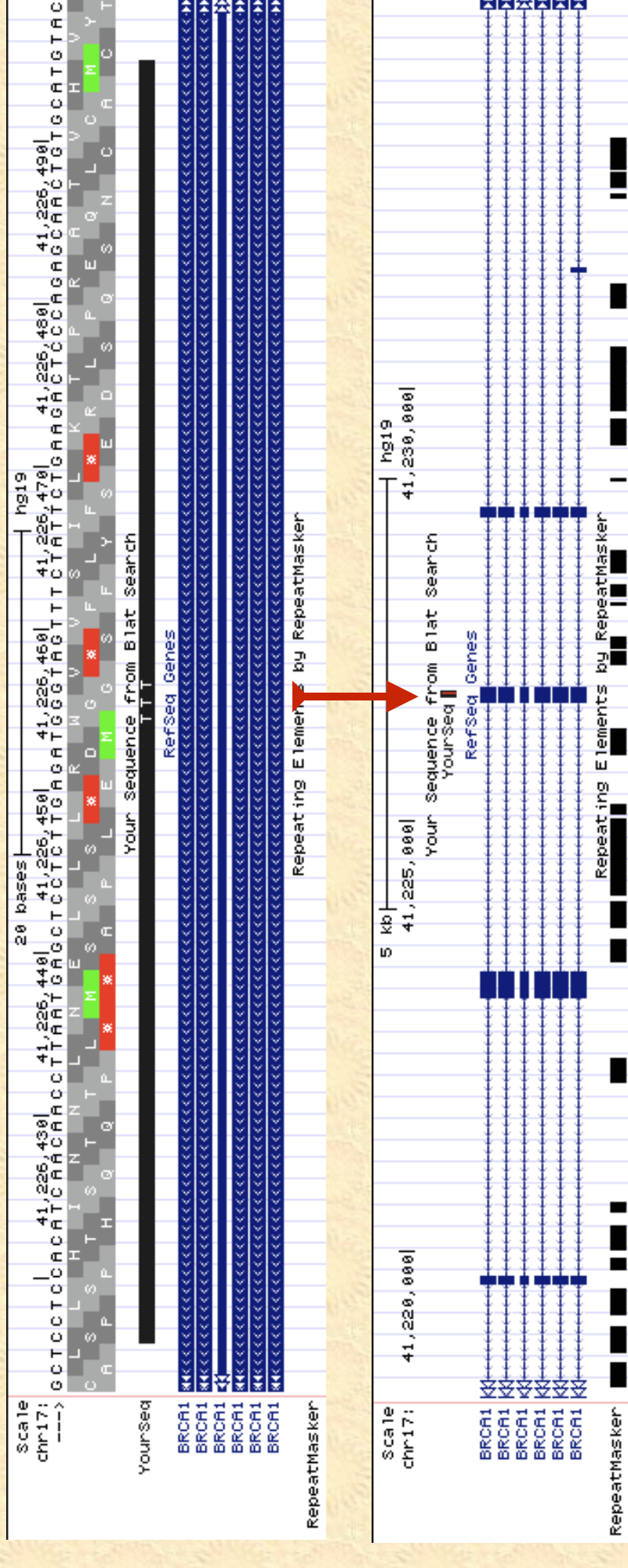


Sequence Search - Blat Results

Human BLAT Results

BLAT Search Results

ACTIONS	QUERY	SCORE	START	END	QSIZE	IDENTITY	CHRO	STRAND	START	END	SPAN
browser details	YourSeq	73	1	79	79	96.3%	17	+	41226417	41226495	79
browser details	YourSeq	25	41	66	79	100.0%	5	+	116751191	116751353	163
browser details	YourSeq	20	30	49	79	100.0%	6	+	44372876	44372895	20
browser details	YourSeq	20	37	56	79	100.0%	11	+	46051786	46051805	20



A Taste of the UCSC Genome Browser

Navigation

Tracks

Sequence
Search
(blat)

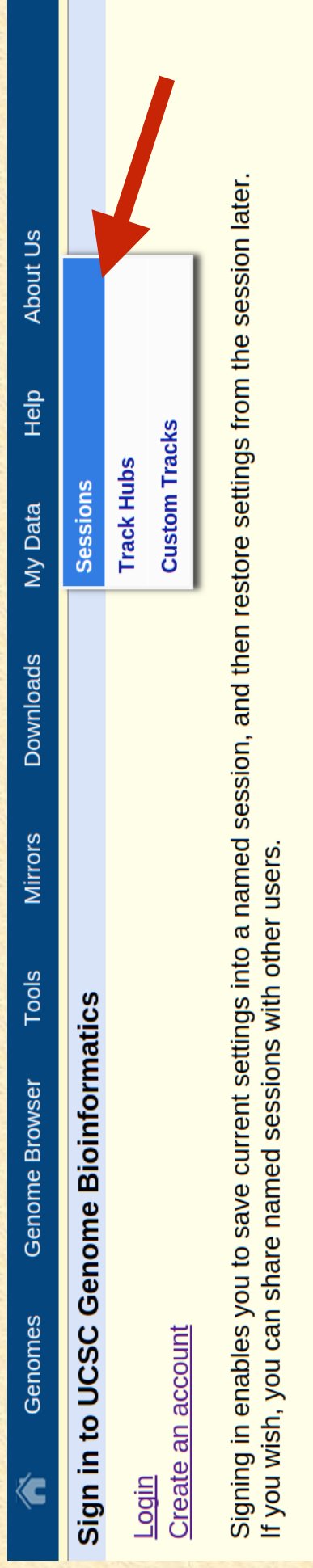
Sessions,
Data Sharing

Table
Browser

Custom
Tracks

Sessions

- Sessions allow:
 - saving the current view
 - sharing it with others
 - loading the saved view at a later time (or a different computer)



The screenshot shows the UCSC Genome Bioinformatics website interface. At the top is a dark blue navigation bar with white text links: Genomes, Genome Browser, Tools, Mirrors, Downloads, My Data, Help, and About Us. Below this bar, on the left, are links for "Sign in to UCSC Genome Bioinformatics", "Login", and "Create an account". The "Sessions" link in the navigation bar is highlighted, and a dropdown menu is open, showing "Sessions", "Track Hubs", and "Custom Tracks". A red arrow points to the "Sessions" option in the dropdown menu. Below the navigation bar, the main content area contains the text: "Signing in enables you to save current settings into a named session, and then restore settings from the session later. If you wish, you can share named sessions with other users."

Sessions

[Home](#) [Genomes](#) [Genome Browser](#) [Tools](#) [Mirrors](#) [Downloads](#) [My Data](#) [Help](#) [About Us](#)

Welcome agordon1

Your Account Information

Username: agordon1
[Change password](#)
[Sign out](#)

Session Management

See the [Sessions User's Guide](#) for more information about this tool.
[Click here to reset](#) the browser user interface settings to their defaults.

My Sessions

session name	created on	use this session	delete this session	share with others?	link to session	send to mail
hg38-my-favorite-gene	2015-03-18	use	delete	☒	Browser	Email

**Restore saved view
(or share with a collaborator)**

Save Settings

Save current settings as named session:
name: ☒ allow this session to be loaded by others

Save current settings to a local file:
file: file type returned:
(leave file blank to get output in browser window)

Save current view

Saving as PDF

The screenshot displays the UCSC Genome Browser interface. The top navigation bar includes links for Genomes, Genome Browser, Tools, Mirrors, Downloads, My Data, View, Help, and About Us. The main title is "UCSC Genome Browser on Human Feb. 2009". Below the title, there are navigation controls: a "move" button with navigation arrows, a "zoom in" button with a magnifying glass, and a "zoom out" button with a minus sign. A text input field shows the current position: "chr17:41,253,535-41,267,057 13,523 bp." Below this, a track for "chr17 (q21.31)" is shown, with a red vertical line indicating the current position. The main display area shows a genomic track with a scale bar and a "RefSeq" track. A red arrow points to the "PDF/PS" menu option in the top navigation bar.

PDF/PS
DNA
In Other Genomes (Convert)
Ensembl
NCBI
Configure Browser
Default Tracks
Default Track Order
Reset All User Settings

**Save current view as PDF
(publication quality)**

Saving as PNG

The screenshot displays the UCSC Genome Browser interface. The top navigation bar includes links for Genomes, Genome Browser, Tools, Mirrors, Downloads, My Data, View, Help, and About Us. The main title is "UCSC Genome Browser on Human Feb. 2009". Below the title, there are navigation controls including a "move" button, a search bar with the text "chr17:41,253,535-41,267,057 13,523 bp.", and a "zoom in" button. The main content area shows a genomic track for chromosome 17 (q21.31) with a scale bar indicating 10 kb. The track displays the BRCA1 gene structure with exons represented by blue bars and introns by lines with arrows. A red arrow points to the "PDF/PS" button in the top navigation bar, which is highlighted in the image.

PDF/PS
DNA
In Other Genomes (Convert)
Ensembl
NCBI
Configure Browser
Default Tracks
Default Track Order
Reset All User Settings

**Save current view as a picture,
can be embedded in a website or email**

A Taste of the UCSC Genome Browser

Navigation

Tracks

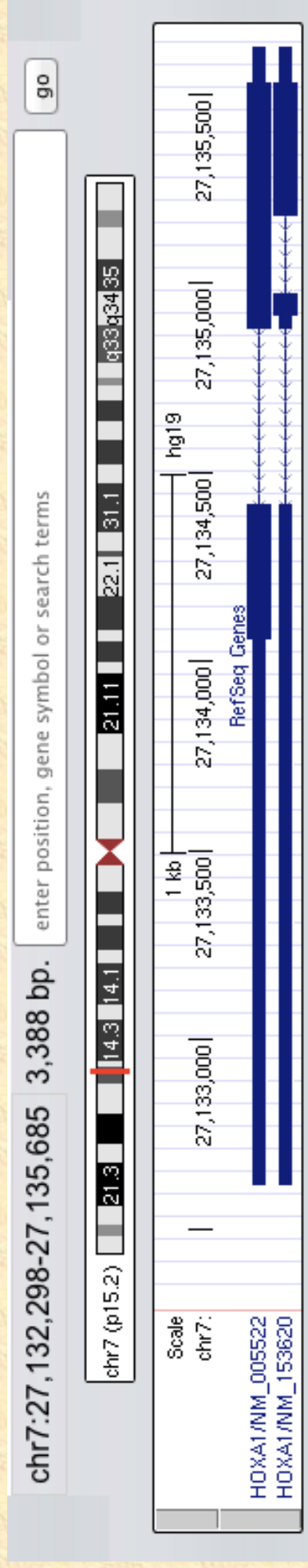
Sequence
Search
(blat)

Sessions,
Data Sharing

Table
Browser

Custom
Tracks

Table Browser - Tracks vs Tables



chr7	27132613	27135625	NM_005522	0	-	27134058	27135531	0	2	1801,746,	0,2266,
chr7	27132613	27135625	NM_153620	0	-	27134914	27135531	0	3	1801,95,448,	0,2266,2564,

Table Browser

Genomes Genome Browser Tools Mirrors Downloads My Data Help About Us

Human (*Homo sapiens*) Genome Browser

group genome

Human Genome Browser – hg19 assembly [sequences]

Blat
Table Browser
Variant Annotation Integrator
Gene Sorter
Genome Graphs
In-Silico PCR
LiftOver
VisiGene
Other Utilities

position
chr12:51,385,091-59,406,178
browser user interface settings to their default
track hubs
configure tracks and display

The February 2009 human reference sequence (GRCh37) was produced by the [Genome Reference Consortium](#). For information about this assembly, see [GRCh37](#) in the NCBI Assembly database.

Table Browser

Table Browser

clade: **genome:** **assembly:**

group: **track:**

table:

region: ☒ genome ☐ position

identifiers (names/accessions):

filter:

intersection:

correlation:

output format:

output file: (leave blank to keep output in browser)

file type returned: ☒ plain text ☐ gzip compressed

To reset **all** user cart settings (including custom tracks), [click here](#).

Table Browser

Table Browser

clade: **genome:** **assembly:**

group:

table:

region: ☒ genome ☐ position

identifiers (names/accessions):

filter:

intersection:

correlation:

output format: [GenomeSpace](#)

output file: (leave blank to keep output in browser)

file type returned: ☒ plain text ☐ gzip compressed

To reset **all** user cart settings (including custom tracks), [click here](#).

Table Browser

Table Browser

clade:

group:

table:

region: ☒ genome ☐ position

identifiers (names/accessions):

filter:

intersection:

correlation:

output format: [GenomeSpace](#)

output file:

file type returned: ☒ plain text ☐ gzip compressed

To reset **all** user cart settings (including custom tracks), [click here](#).

Table Browser

Table Browser

clade: **genome:** **assembly:**

group: **track:**

table:

region: ☒ genome ☐ position

identifiers (names/accessions):

filter:

intersection:

correlation:

output format: [GenomeSpace](#)

output file: (leave blank to keep output in browser)

file type returned: ☒ plain text ☐ gzip compressed

To reset **all** user cart settings (including custom tracks), [click here](#).

Table Browser

Table Browser

clade: **genome:** **assembly:**

group: **track:**

table:

region: ☒ genome ☐ position

identifiers (names/accessions):

filter:

intersection:

correlation:

output format:

output file: (leave blank to keep output in browser)

file type returned: ☒ plain text ☐ gzip compressed

"BED" - list of features
in a simple format.

To reset **all** user cart settings (including custom tracks), [click here](#).

Table Browser

Table Browser

clade: **genome:** **assembly:**

group: **track:**

table:

region: ☒ genome ☐ position

identifiers (names/accessions):

filter:

intersection:

correlation:

output format:

output file: (leave blank to keep output in browser)

file type returned: ☒ plain text ☐ gzip compressed

To reset **all** user cart settings (including custom tracks), [click here](#).

Table Browser - BED options

Output knownGene as BED

☐ Include [custom track header](#):

name=

description=

visibility=

url=

Create one BED record per:

- ☒ Whole Gene
- ☐ Upstream by bases
- ☐ Exons plus bases at each end
- ☐ Introns plus bases at each end
- ☐ 5' UTR Exons
- ☐ Coding Exons
- ☐ 3' UTR Exons
- ☐ Downstream by bases

Note: if a feature is close to the beginning or end of a chromosome and upstream/downstream bases are added, they may be truncated in order to avoid extending past the edge of the chromosome.

Table Browser - BED output

chr1	11873	14409	uc001aaa.30	+	11873118730	3	354,109,1189,	0,739,1:
chr1	11873	14409	uc001aab.30	+	11873118730	3	354,52,1189,	0,772,1347
chr1	11873	14409	uc010nxq.10	+	12189136390	3	354,127,1007,	0,721,1:
chr1	17368	17436	uc031tla.10	-	17368173680	1	68,	0,
chr1	14361	19759	uc001aae.40	-	14361143610	10	468,69,152,159,198,136,	
chr1	14361	19759	uc009vit.30	-	14361143610	9	468,69,152,159,198,510,	
chr1	14361	19759	uc009viu.30	-	14361143610	10	468,69,152,159,198,510,	
chr1	16857	19759	uc001aai.10	-	16857168570	6	198,136,137,147,112,847	
chr1	14361	29370	uc001aah.40	-	14361143610	11	468,69,152,159,198,136,	
chr1	14361	29370	uc009viq.30	-	14361143610	7	468,152,159,198,456,154	
chr1	14406	29370	uc009viv.20	-	14406144060	7	2359,198,136,137,147,15	
chr1	16606	29370	uc009viz.20	-	16606166060	8	159,202,136,137,147,112	
chr1	16606	29370	uc009viy.20	-	16606166060	8	159,198,136,137,147,95,	
chr1	17232	29370	uc009vje.20	-	17232172320	4	510,147,99,50,	0,682,10
chr1	17605	29370	uc009vjf.20	-	17605176050	7	137,147,95,58,227,154,5	
chr1	16857	29961	uc009vjb.10	-	16857168570	7	198,136,137,147,112,154	
chr1	30365	30503	uc031tlb.10	+	30365303650	1	138,	0,
chr1	34610	36081	uc001aak.30	-	34610346100	3	564,205,361,	0,666,1110
chr1	69090	70008	uc001aal.10	+	69090700080	1	918,	0,

Table Browser - Intersections

Table Browser

clade: **genome:** **assembly:**

group: **track:**

table:

region: ☒ genome ☐ position

identifiers (names/accessions):

filter:

intersection:

correlation:

output format:

output file: (leave blank to keep output in browser)

file type returned: ☒ plain text ☐ gzip compressed

To reset **all** user cart settings (including custom tracks), [click here](#).

A Taste of the UCSC Genome Browser

Navigation

Tracks

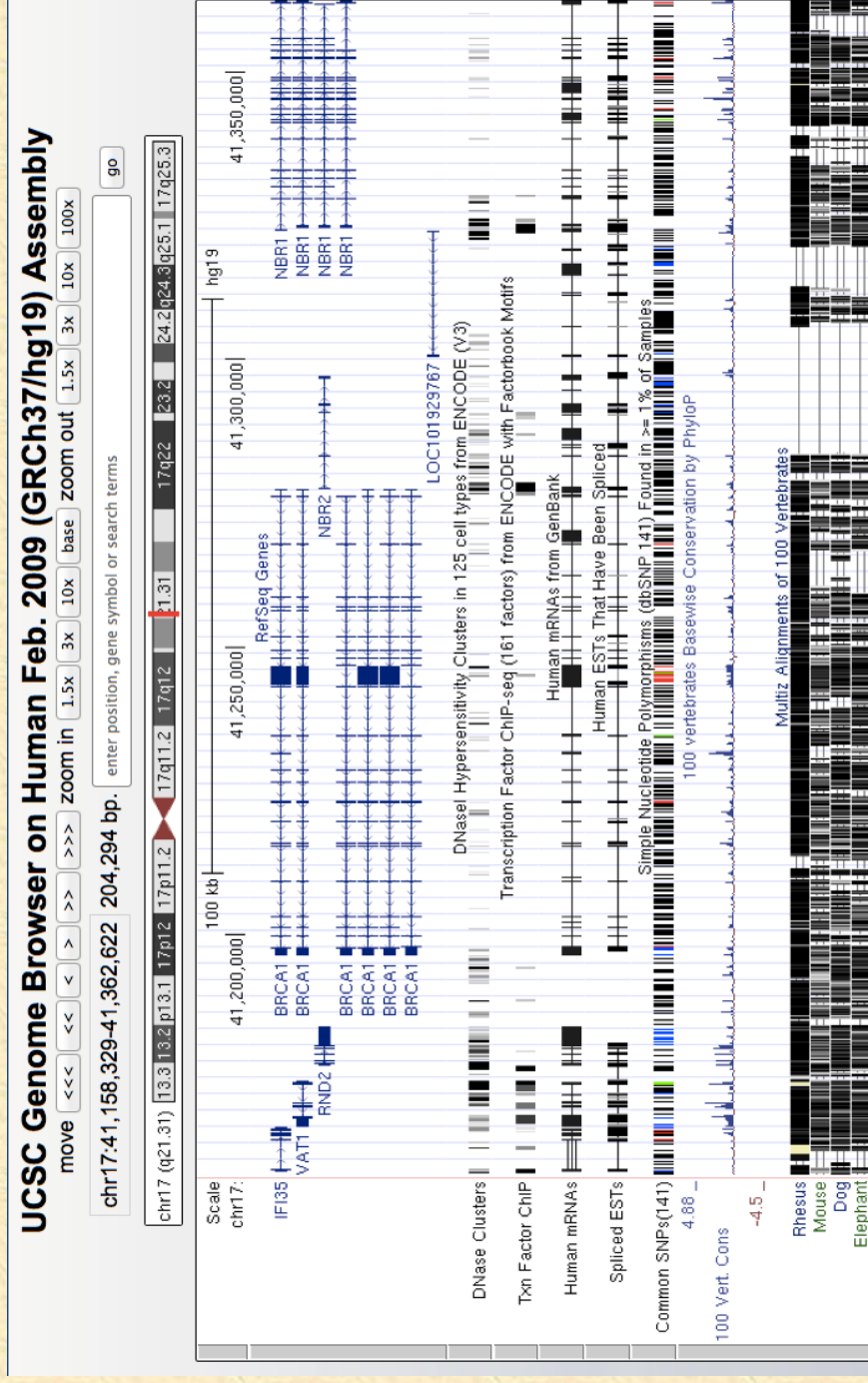
Sequence
Search
(blat)

Sessions,
Data Sharing

Table
Browser

Custom
Tracks

Custom Tracks



Click on a feature for details. Click or drag in the base position track to zoom in. Click side bars for track options. Drag side bars or labels up or down to reorder tracks. Drag tracks left or right to new position.

track search

default tracks

default order

hide

add custom tracks

track hubs

configure

reverse

resize

refresh

collapse all

Use drop-down controls below and press refresh to alter tracks displayed.
Tracks with lots of items will automatically be displayed in more compact modes.

expand all

Custom Tracks - BED3

Add Custom Tracks

clade genome assembly [dm3]

Paste URLs or data:

```
chr2L 3000 3032
chr2L 3020 3049
chr2L 3028 3063
chr2L 3054 3078
```

Data can be as simple as chromosome, start, end (BED3 format)

Custom Tracks - BED3

Add Custom Tracks

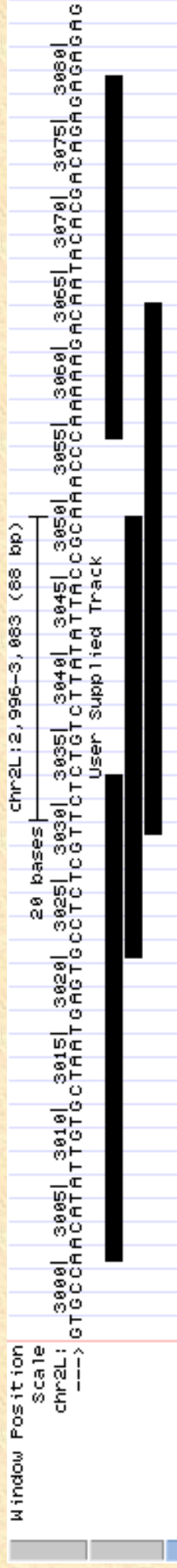
clade genome assembly [dm3]

Paste URLs or data:

Or upload:

```
chr2L 3000 3032
chr2L 3020 3049
chr2L 3028 3063
chr2L 3054 3078
```

**Data can be as simple as
chromosome, start, end
(BED3 format)**



Custom Tracks - BED6

Add Custom Tracks

clade

Insect

genome

D. melanogaster

assembly

Apr. 2006

[dm3]

Paste URLs or data:

Or upload:

Submit

Clear

chr2L	3000	3032	Inigo	100	+
chr2L	3020	3049	Westley	400	+
chr2L	3028	3063	Vizzini	700	-
chr2L	3054	3078	Fezzik	1000	+

**chromosome, start, end, name, score (color), strand
(BED6 format)**



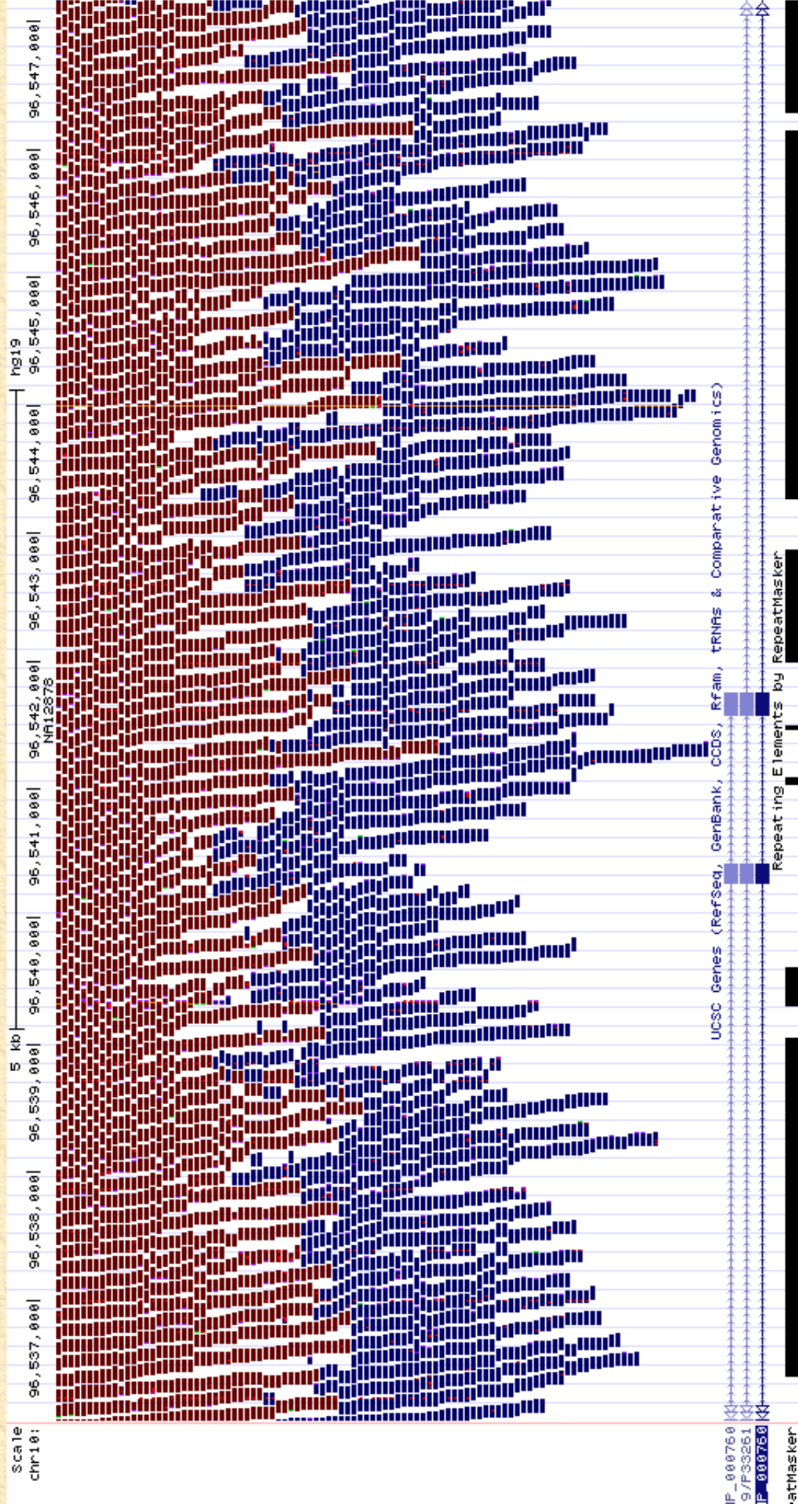
Custom Tracks - 'Big' tracks

- High throughput sequencing generates very large files
 - too large to be uploaded (or copy&pasted)
- Custom tracks can be added by storing the files on a public server (e.g. institute's servers), and pointing to UCSC Genome Browser to the public location (URL):

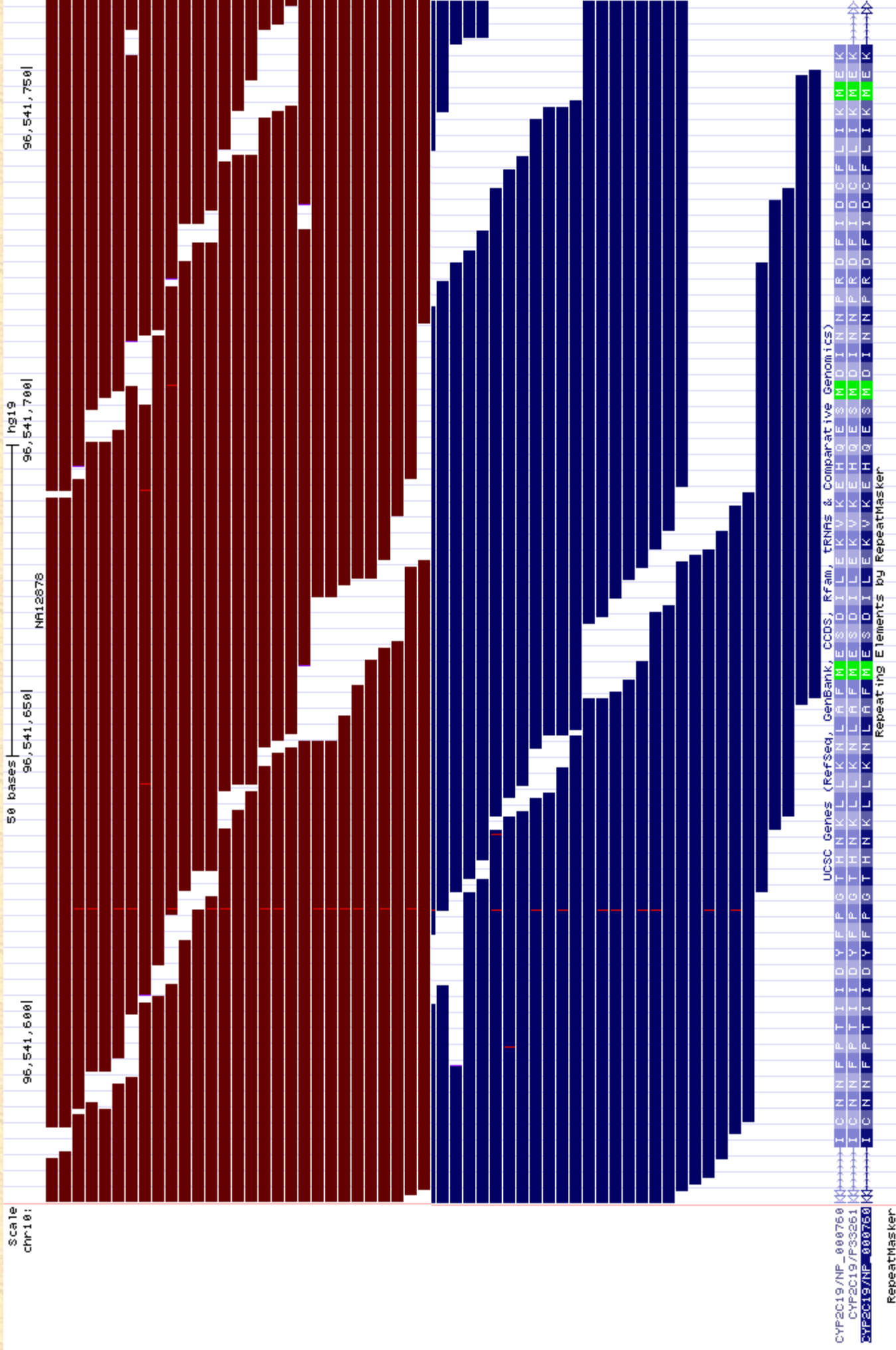
Add Custom Tracks

clade	Mammal	genome	Human	assembly	Feb. 2009 (GRCh37/hg19)
Paste URLs or data:	Or upload:		Choose File	no file selected	
track type=bam name=NA12878 bigDataUrl=http://ampx-smash-sampled-human-NA12878.s3.amazonaws.com/NA12878.bam					
					Submit
					Clear

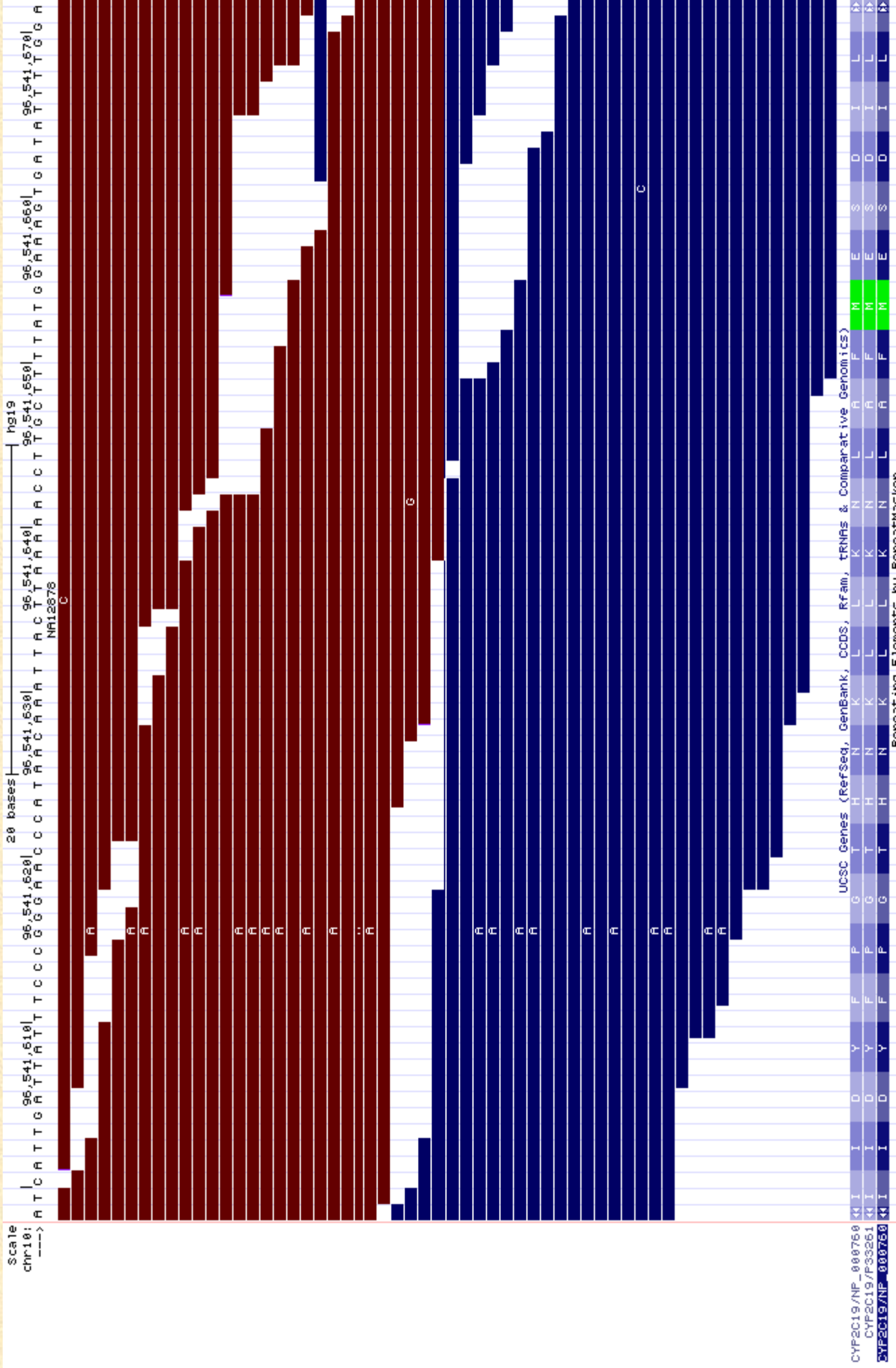
Custom Tracks - 'Big' Tracks (BAM)



Custom Tracks - 'Big' Tracks (BAM)

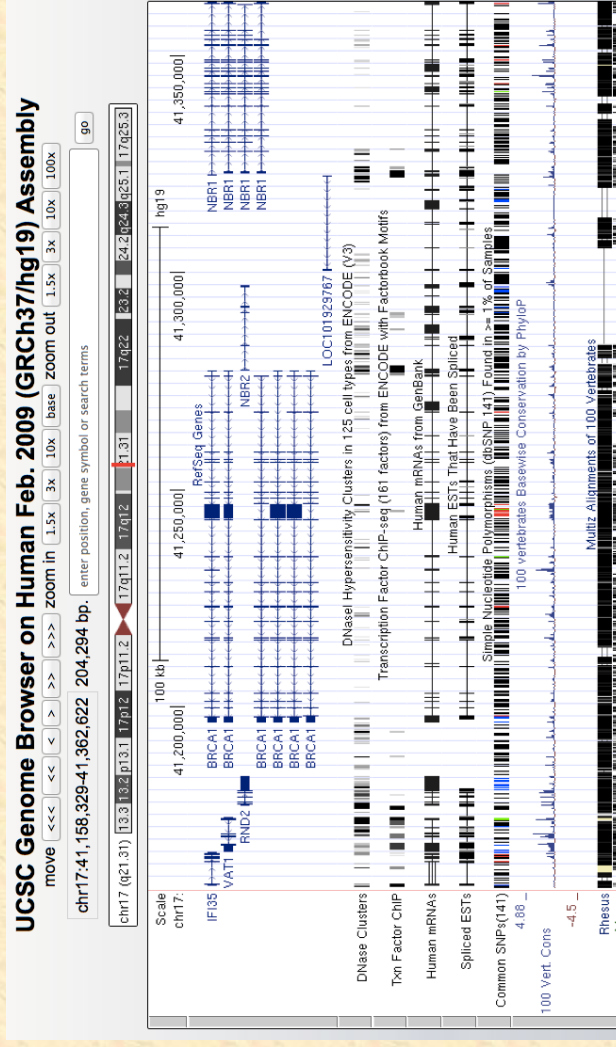


Custom Tracks - 'Big' Tracks (BAM)



Custom Tracks - Hubs

- A collection of custom tracks can be published as one group
- Suitable for supplementary data
- Easier for users



Click on a feature for details. Click or drag in the base position track to zoom in. Click side bars for track options. Drag side bars or labels up or down to reorder tracks. Drag tracks left or right to new position.

Use drop-down controls below and press refresh to alter tracks displayed.

Tracks with lots of items will automatically be displayed in more compact modes.

Custom Tracks - Hubs

Track Data Hubs

Public Hubs

My Hubs

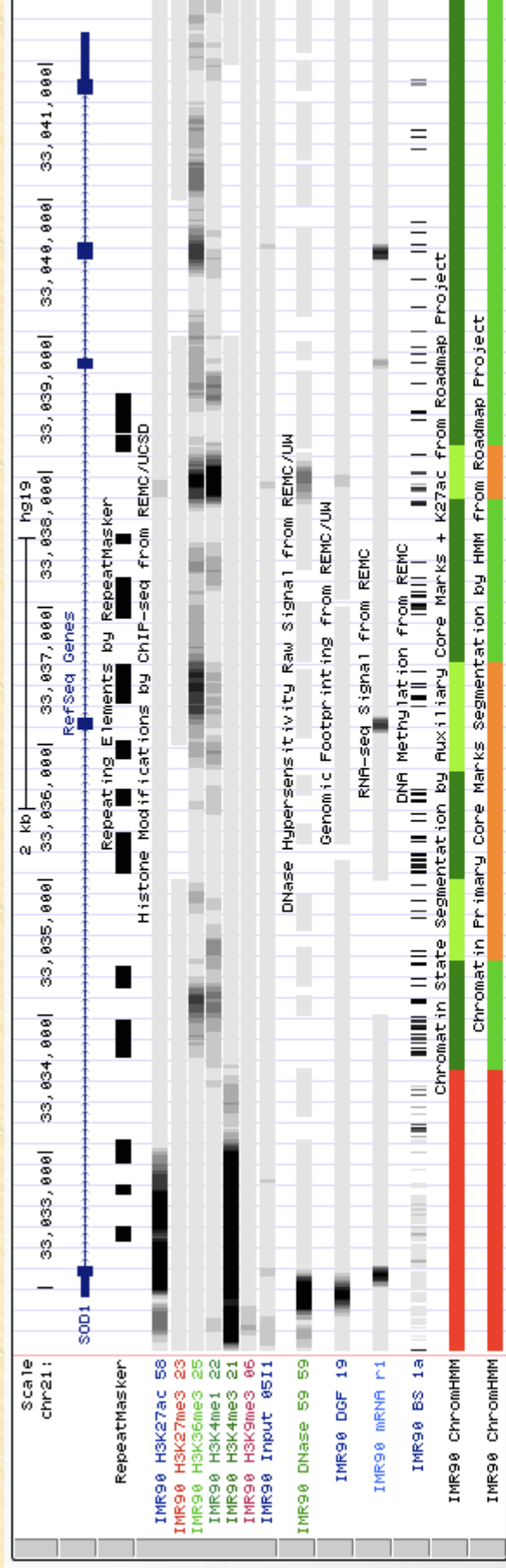
Enter search terms to find in public track hub description pages:

Search Public Hubs

Clicking Connect redirects to the gateway page of the selected hub's default assembly.

Display	Hub Name	Description	Assemblies
Connect	Roadmap Epigenomics Data Complete Collection at Wash U VizHub	Roadmap Epigenomics Human Epigenome Atlas Data Complete Collection, VizHub at Washington University in St. Louis	hg19
Connect	UMassMed ZHub	UMassMed H3K4me3 ChIP-seq data for Autistic brains	hg19
Connect	Cancer genome polyA site & usage	An in-depth map of polyadenylation sites in cancer (matched-pair tissues and cell lines)	hg19
Connect	ENCODE Analysis Hub	ENCODE Integrative Analysis Data Hub	hg19
Connect	miRcode microRNA sites	Predicted microRNA target sites in GENCODE transcripts	hg19
Connect	Translation Initiation Sites (TIS)	Translation Initiation Sites (TIS) track	hg19

Tracks - Hubs



move start move end

Click on a feature for details. Click or drag in the base position track to zoom in. Click side bars for track options. Drag side bars or labels up or down to reorder tracks. Drag tracks left or right to new position.

track search

Use drop-down controls below and press refresh to alter tracks displayed. Tracks with lots of items will automatically be displayed in more compact modes.

Roadmap Epigenomics Data Complete Collection at Wash U VizHub

[Broad Histone](#) [UCSD Histone](#)

[DNA Methylation](#)

[Roadmap ChromHMM](#)

[ChromHMM](#)

[Auxiliary Core Marks](#)

[UCSF-UBC-USC](#)

[Histone](#)

[By Assay...](#)

[Footprinting](#)

[Sample Summary...](#)

[RNA](#)

[Methylation Summary...](#)

[Roadmap ChromHMM](#)

[Primary Core Marks](#)

[Roadmap Uniformly Signal](#)

A Taste of the UCSC Genome Browser

Navigation

Tracks

Sequence
Search
(blat)

Sessions,
Data Sharing

Table
Browser

Custom
Tracks

Genome Browser - Help!

- Explore! Experiment!
 - you can't do any damage
- UCSC Genome Browser - User Guide
 - <http://genome.ucsc.edu/goldenpath/help/hgTracksHelp.html>
 - FAQ: <http://genome.ucsc.edu/FAQ/>
 - Training: <http://genome.ucsc.edu/training/index.html>
- Free UCSC Genome Browser Video Tutorials by OpenHelix:
 - Introduction: <http://www.openhelix.com/tutorialInfo.cgi?id=27>
 - Custom Track and Table Browser: <http://www.openhelix.com/tutorialInfo.cgi?id=28>
 - Additional Tools: <http://www.openhelix.com/tutorialInfo.cgi?id=76>
- Ask questions by email: genome@soe.ucsc.edu
- Search past questions/answers archives:
<https://groups.google.com/a/soe.ucsc.edu/forum/#!forum/genome>
- Ask questions on BioStar: <https://www.biostars.org>
- UCSC Workshop at your institution

Thank You!

Questions?

Appendix - NA12878

- Subject NA12878

https://catalog.coriell.org/0/sections/search/Sample_Detail.aspx?Ref=NA12878&product=DNA

"...Mother; donor subject has a single bp (G-to-A) transition at nucleotide 681 in exon 5 of the CYP2C19 gene (CYP2C19*2) which creates an aberrant splice site."

- 1000 Genome Project BAM Files:

<ftp://ftp.1000genomes.ebi.ac.uk/vol1/ftp/data/NA12878/alignment/>

- High-Coverage BAM File (114GB):

<http://smash.cs.berkeley.edu/datasets.html>

- Low-Coverage UCSC Custom Track Line:

[track type=bam name=NA12878 bigDataUrl=ftp://ftp.1000genomes.ebi.ac.uk/vol1/ftp/data/NA12878/alignment/NA12878.mapped.ILLUMINA.bwa.CEU.low_coverage.20121211.bam](http://ftp.1000genomes.ebi.ac.uk/vol1/ftp/data/NA12878/alignment/NA12878.mapped.ILLUMINA.bwa.CEU.low_coverage.20121211.bam)

- High-Coverage UCSC Custom Track Line:

[track type=bam name=NA12878 bigDataUrl=http://ampx-smash-sampled-human-NA12878.s3.amazonaws.com/NA12878.bam](http://ampx-smash-sampled-human-NA12878.s3.amazonaws.com/NA12878.bam)

- CYP2C19 Exon 5 (G->A SNP): chr10:96,541,611-96,541,620

The Genome Access Course

Ensembl Genome Browser





Ensembl

Search: All species for e.g. BRCA2 or rat X:100000..200000

Browse a Genome

The Ensembl project produces genome databases for vertebrates and other eukaryotic species, and makes this information freely available online.

★ Log in to customize this list

All genomes

View full list of all Ensembl species

Other species are available in Ensembl Pro and Ensembl Genomes

What's New in Release 71 (April 2013)

- New Expression view listing RNAseq data
- New assemblies for Chicken (Galgal4) and C. elegans (WBcel235)
- Scrollable image has been integrated into Region in Detail

Full details of this release

More release news on our blog →

Did you know...?

Have some data? View it in Ensembl by uploading a file or attaching a url.

The Genome Access Course

Search: for

e.g. BRCA2 or rat X:100000..200000 or coronary heart disease

Browse a Genome

The Ensembl project produces genome databases for vertebrates and other eukaryotic species, and makes this information freely available online.

Popular genomes

Human GRCh37 **Mouse** GRCm38 **Zebrafish** Zv9

★ [Log in to customize this list](#)

All genomes

[View full list of all Ensembl species](#)

Other species are available in [Ensembl Pre!](#) and [EnsemblGenomes](#)

ENCODE data in Ensembl

Variant Effect Predictor

Gene expression in different tissues

Find SNPs and other variants for my gene

Retrieve gene sequence

Compare genes across species

Use my own data in Ensembl

Learn about a disease or phenotype

April 2015

The Genome Access Course

EnsemblGenomes

e!EnsemblGenomes

About us | Genomes | Data types | Downloads

Bacteria | Protists | Fungi | Plants | Metazoa | Vertebrates

e!EnsemblProtists ▾ BLAST | Sequence Search | BioMart | Tools | Downloads | Help | Documentation

e!EnsemblFungi ▾ BLAST | Sequence Search | BioMart | Tools | Downloads | Help | Documentation

e!EnsemblPlants ▾ BLAST | Sequence Search | BioMart | Tools | Downloads | Help | Documentation

e!EnsemblMetazoa ▾ BLAST | Sequence Search | BioMart | Tools | Downloads | Help | Documentation

e!EnsemblBacteria ▾ Sequence Search | BLAST | Tools | Downloads | Help | Documentation

e!Ensembl BLAST/BLAT | BioMart | Tools | Downloads | Help & Documentation | Blog | Mirrors

Ensembl Vertebrates is the default Ensembl view

April 2015

Ensembl is organized by organism genome: the primary website features eukaryotic genomes



Ensembl Species

Alpaca <i>Vicugna pacos</i> vicPac1	Fugu <i>Tetraodon rubripes</i> FUGU4	Pika <i>Ochotona princeps</i> pika
Anole lizard <i>Anolis carolinensis</i> AnoCar2.0	Gibbon <i>Nomascus leucogenys</i> Nleu1.0	Platyfish <i>Xiphophorus maculatus</i> Xipmac4.4.2
Armadillo <i>Dasypus novemcinctus</i> dasNov2	Gorilla <i>Gorilla gorilla gorilla</i> gorGor3.1	Platypus <i>Ornithorhynchus anatinus</i> OANAS
Baboon (preview - assembly only) <i>Papio hamadryas</i> Pham	Guinea Pig <i>Cavia porcellus</i> cavPor3	Rabbit <i>Oryctolagus cuniculus</i> oryCun2
Budgerigar (preview - assembly only) <i>Melopsittacus undulatus</i> MelUnd6.3	Hedgehog <i>Erinaceus europaeus</i> HEDGEHOG	Rat <i>Rattus norvegicus</i> Rnor_5.0
Bushbaby <i>Otolemur garnettii</i> OtoGar3	Horse <i>Equus caballus</i> EquCab2	Saccharomyces cerevisiae <i>Saccharomyces cerevisiae</i> EF4
Ciona intestinalis <i>Ciona intestinalis</i> KH1	Human <i>Homo sapiens</i> GRCh37	Sheep (preview - assembly only) <i>Ovis aries</i> oviAri1
Ciona savignyi <i>Ciona savignyi</i> CSAV2.0	Hyrax <i>Procavia capensis</i> proCap1	Shrew <i>Sorex araneus</i> COMMON_SHREW1
Caenorhabditis elegans <i>Caenorhabditis elegans</i> WBcel235	Kangaroo rat <i>Dipodomys ordii</i> dipOrd1	Sloth <i>Choloepus hoffmanni</i> choHof1

April 2015

Ensembl has two other sites: Pre!Ensembl (in-process genomes) and EnsemblGenomes (extends Ensembl across taxa)



About Pre Ensembl

The Ensembl pre-build site provides displays of genomes that are in the process of being annotated. Not all standard Ensembl displays are available. Genomes are put up here when we have done the initial BLAST analysis on a new assembly but have not completed the gene build. It is provided as an "early access" site for our users. Owing to the preliminary nature of the data, Pre-Ensembl provides views of the assembly, BLAST against the assembly and download of portions of the assembly - and little else. In general a full Ensembl release takes months depending on how complex the data are and the time constraints of people in the team.

Browse a genome

Baboon <i>Papio</i>	Flycatcher <i>PicAp_1.4</i> NEW!
Budgerigar <i>MelUnd6.3</i>	Painted Turtle <i>ChrPicBel3.0.1</i> NEW!
Chicken <i>Gallus</i>	Pig FPC map [clone map only; the full genome annotation on the SScrofa3 assembly can be found here]
Chimpanzee <i>Pan_troglodytes_2.1.3</i>	Sheep <i>oviAri1</i>

Pre!



Genomes

Ensembl Genomes: Extending Ensembl across the taxonomic space.



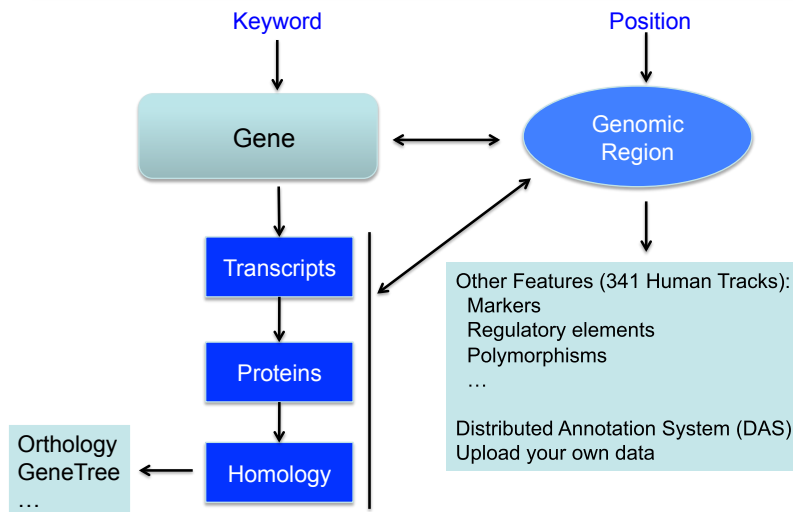
in EnsemblBacteria

Over 6000 bacterial genome sequences have been annotated and deposited in the public archives of the members of the [International Nucleotide Sequence Database Collaboration](#). This site provides access to complete, annotated bacterial genomes (present in the [European Nucleotide Archive](#)) through the Ensembl graphical user interface ([genome browser](#)). More details about the integration are

April 2015

Each Ensembl genome features genome browsing, sequence comparison, and bulk querying via BioMart

Querying Ensembl from different perspectives returns related data sets



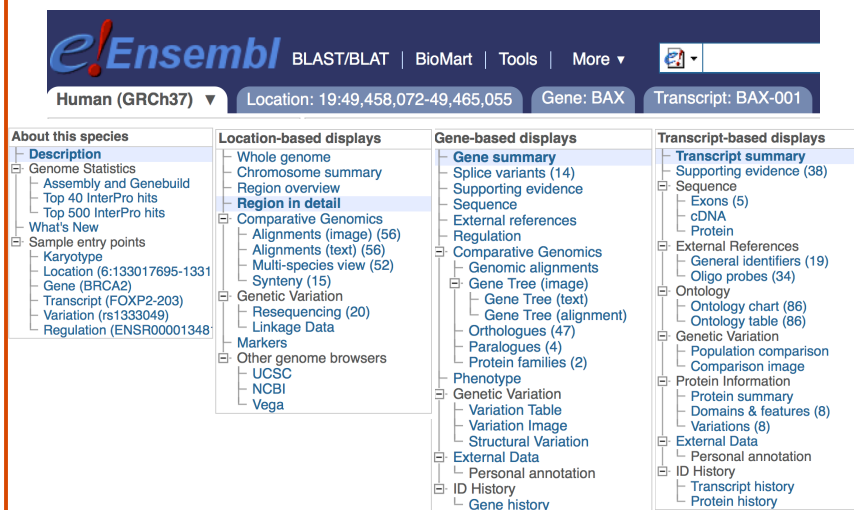
For most queries, there are four steps to get access a large amount of information in Ensembl

1. Select species
2. Select gene
3. View Transcript
4. Click tabs to access the data



April 2015

Selecting the many options will lead to expanded directories of information



April 2015

Ensembl uses stable identifiers for genes, transcripts, proteins, exons, and regulatory features

Human:

- ENSG##### Gene ID
- ENST##### Transcript ID
- ENSP##### Protein ID
- ENSE##### Exon ID
- ENSR##### Regulatory Feature ID

Human (GRCh37) Location: 19:49,458,072-49,465,055 Gene: BAX

Gene-based displays

- Gene summary
- Splice variants (13)
- Transcript comparison
- Supporting evidence
- Sequence
- External references
- Regulation
- Expression
- Comparative Genomics

Gene: BAX ENSG00000087088

Description BCL2-associated X protein [Source:HGNC Symbol;Acc:959]

Location [Chromosome 19: 49,458,072-49,465,055](#) forward strand.

INSDC coordinates chromosome:GRCh37:CM000681.1:49458072:49465055:1

Transcripts This gene has 13 transcripts (splice variants) [Show transcript table](#)

Other species contain an affix derived from their Latin name:

- ENSMUSG##### A mouse **gene** (*Mus*)
- ENSDDART##### A zebrafish **transcript** (*Danio rerio*)

April 2015

Ensembl uses a tiered GeneBuild system: automated and manual (Vega/Havana) curation merged when identical

Chromosome bands

Genes (GENCODE...)

Havana/manual

HGNC Symbol: BAX-006

Transcript ENS000000345358

Protein ENSP00000263262

Gene ENSG00000087088

Location [Chromosome 19: 49,458,134-49,465,055](#)

Gene type Known protein coding

Transcript type Known protein coding

Strand Forward

Base pairs 793

Amino acids 192

Analysis Ensembl/Havana merge

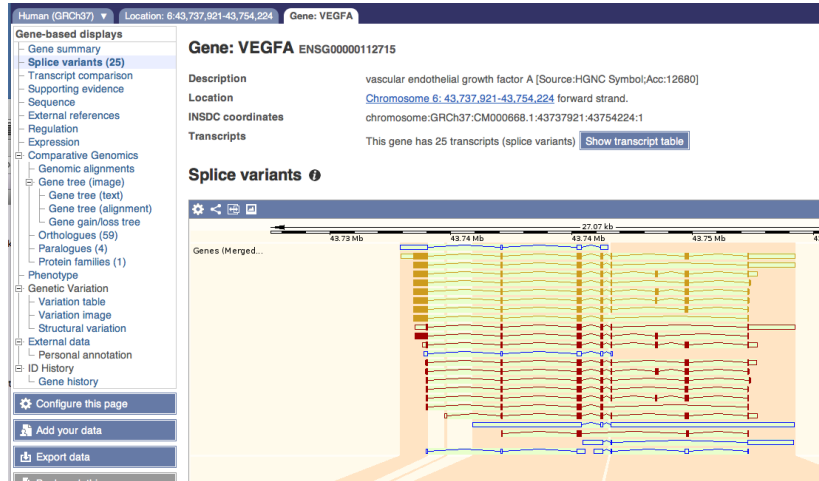
Prediction Transcript where the method

Ensembl genebuild transcript and the Vega manual annotation have the same sequence, for every base pair. See [article](#).

automated

April 2015

The Ensembl tiered GeneBuild system will also display all known splice variants



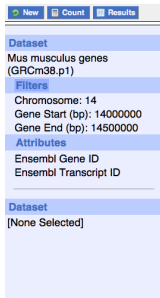
April 2015

Worked example

Browsing Ensembl

April 2015

Data retrieval with Biomart



“Biomart” allows you to export a lot of data in one go in an organized fashion.

Three (easy) steps:

1. Choose dataset (human, mouse etc): **DATASET**
2. Choose bits of data you need (genes, regions of genome, GO term, etc): **FILTER**
3. Choose annotations and format to retrieve: **ATTRIBUTES**

April 2015

Worked example #2

Downloading data using Ensembl

April 2015

Worked example #1 Browsing Ensembl

e!Ensembl BLAST/BLAT | BioMart | Tools | Downloads | More ▾ Login · Register

Search: for

e.g. **BRCA2** or **rat X:100000..200000** or **coronary heart disease**

Browse a Genome

The Ensembl project produces genome databases for vertebrates and other eukaryotic species, and makes this information freely available online.

Click on a link below to go to the species' home page.

Popular genomes ([Log in to customize this list](#))

-  **Human**
GRCh37
-  **Mouse**
NCBIM37
-  **Zebrafish**
Zv9

New to Ensembl?

Did you know you can:

-  [Learn how to use Ensembl](#)
with our video tutorials and walk-throughs
-  [Add custom tracks](#)
using our new Control Panel
-  [Upload and analyse your data](#)
and save it to your Ensembl account
-  [Fetch only the data you want](#)
from our public database, using the Perl API
-  [Download our databases via FTP](#)
in FASTA, MySQL and other formats
-  [Mine Ensembl with BioMart](#)
and export sequences or tables in text, html, or Excel format

Did you know...?
 New to Ensembl? Try our [online course](#).

[Protein sequence](#)

Choose human



Human
Homo sapiens

Search all categories ▾

Search 'BAX'

Only searching Human ▾ **BAX**

839 results match **BAX** when restricted to species: Human ✕

BAX (Human Gene)

ENSG00000087088 19:48954815-48961798:1
BCL2-associated X protein [Source:HGNC Symbol;Acc:HGNC:959]

BAX (Vega gene) is associated with Gene ENSG00000087088
[Variation table](#) • [Location](#) • [Regulation](#) • [Orthologues](#) • [Gene tree](#)

BAX-006 (Human Transcript)

ENST00000345358 19:48954877-48961798:1
BCL2-associated X protein [Source:HGNC Symbol;Acc:HGNC:959]

BAX-006 (Vega transcript) is associated with Transcript ENST00000345358
[Location](#) • [cDNA seq.](#) • [Variation table](#) • [Protein seq.](#) • [Population](#) • [Protein](#)

Click BAX (Human Gene)

Three of the four tabs are available

e!Ensembl BLAST/BLAT | BioMart | Tools | Downloads | Help & Documentation | Blog | Mirrors

Human (GRCh38) Location: 19:48,954,815-48,961,798 Gene: BAX

Gene-based displays

- Summary
- Splice variants (13)
- Transcript comparison
- Supporting evidence
- Alt. alleles
- Sequence
 - Secondary Structure
 - External references
 - Regulation
- Comparative Genomics
 - Genomic alignments
 - Gene tree (image)
 - Gene tree (text)
 - Gene tree (alignment)
 - Gene gain/loss tree
- Orthologues (56)
- Paralogues
- Protein families (2)
- Phenotype
- Genetic Variation
 - Variation table
 - Variation image
 - Structural variation
- External data
 - Personal annotation
- ID History
 - Gene history

Gene: BAX ENSG00000087088

Description BCL2-associated X protein [Source:HGNC Symbol;Acc:HGNC:959]

Synonyms BCL2L4

Location [Chromosome 19: 48,954,815-48,961,798](#) forward strand.

INSDC coordinates chromosome:GRCh38:CM000681.2:48954815:48961798:1

Transcripts

This gene has 13 transcripts (splice variants) [Hide transcript table](#)

Name	Transcript ID	bp	Protein	Biotype	CCDS	RefSeq	Flags
BAX-001	ENST00000293288	1358	218 aa	Protein coding	CCDS12744	NM_004324 NP_001278358 NP_001278359 NP_001278360	TSL:1 GENCODE basic
BAX-006	ENST00000345356	793	192 aa	Protein coding	CCDS12742	NM_138764 NP_001278358 NP_001278359 NP_001278360	TSL:1 GENCODE basic
BAX-003	ENST00000415969	540	179 aa	Protein coding	CCDS12745	NM_138763 NP_001278360 NP_001278361	TSL:1 GENCODE basic
BAX-005	ENST00000354470	433	143 aa	Protein coding	CCDS12743	NM_138763 NP_001278360 NP_001278361	TSL:1 GENCODE basic
BAX-201	ENST00000391871	687	175 aa	Protein coding	-	NP_001278360	TSL:5 GENCODE basic
BAX-202	ENST00000539787	458	140 aa	Protein coding	-	NP_001278359	TSL:1 GENCODE basic
BAX-004	ENST00000506183	383	126 aa	Protein coding	-	NP_001278360	CDS 5' incomplete TSL:1
BAX-002	ENST00000356483	677	164 aa	Nonstop mediated decay	-	-	TSL:1
BAX-007	ENST00000515540	300	41 aa	Nonstop mediated decay	-	-	TSL:3
BAX-009	ENST00000502487	1346	No protein	Retained intron	-	-	TSL:1

The GENCODE set is the gene set for human and mouse. GENCODE Basic is a subset of representative transcripts (splice variants).

Select a transcript to open up the fourth tab. Most confidently annotated transcripts have a 'CCDS' ID.

Summary

6.87 kb Forward strand

Statistics Exons: 5 Coding exons: 5 Transcript length: 1,358 bps Translation length: 218 residues

CCDS This transcript is a member of the Human CCDS set: [CCDS12744](#)

Uniprot This transcript corresponds to the following Uniprot identifiers: [Q07812](#)

Transcript Support Level 1

Ensembl version ENST00000293288.10

Type Known protein coding

Prediction Method Transcript where the Ensembl genebuild transcript and the [Vega](#) manual annotation have the same sequence, for every base pair. See [article](#).

Alternative transcripts This transcript corresponds to the following database identifiers:
Havana transcript: [OTTHUMT00000360762](#)

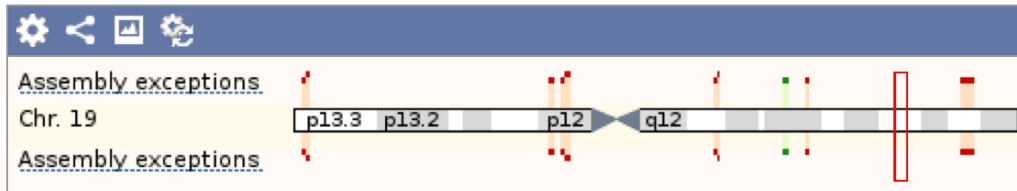
GENCODE basic gene This transcript is a member of the Gencode basic gene set.

Transcript info

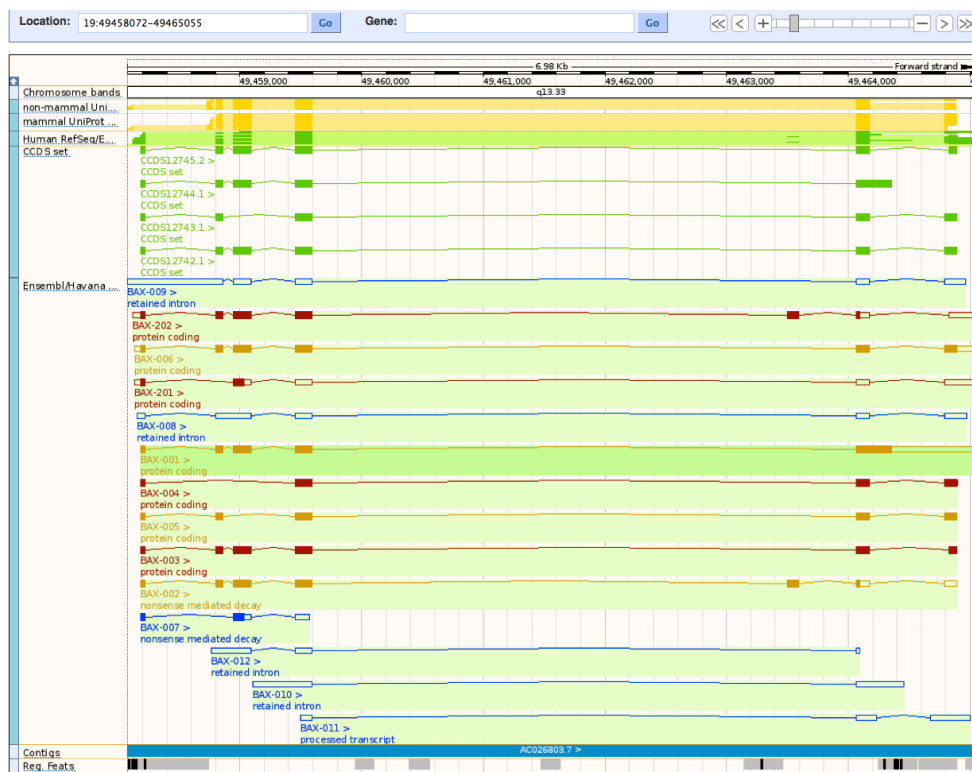
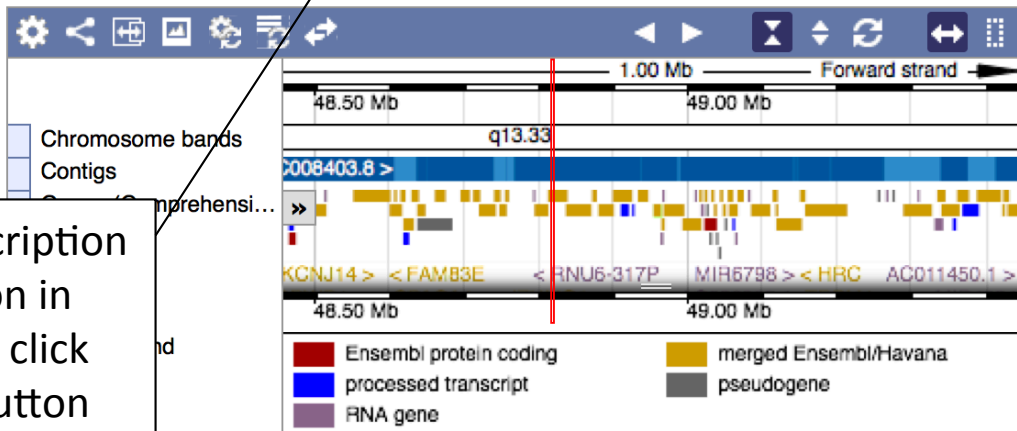
In this module we will focus on viewing the gene in the context of the genome. Future models will focus on other aspects of Ensembl.

To view BAX on the genome – select the 'location' tab

Chromosome 19: 48,954,815-48,961,798



Region in detail ⓘ

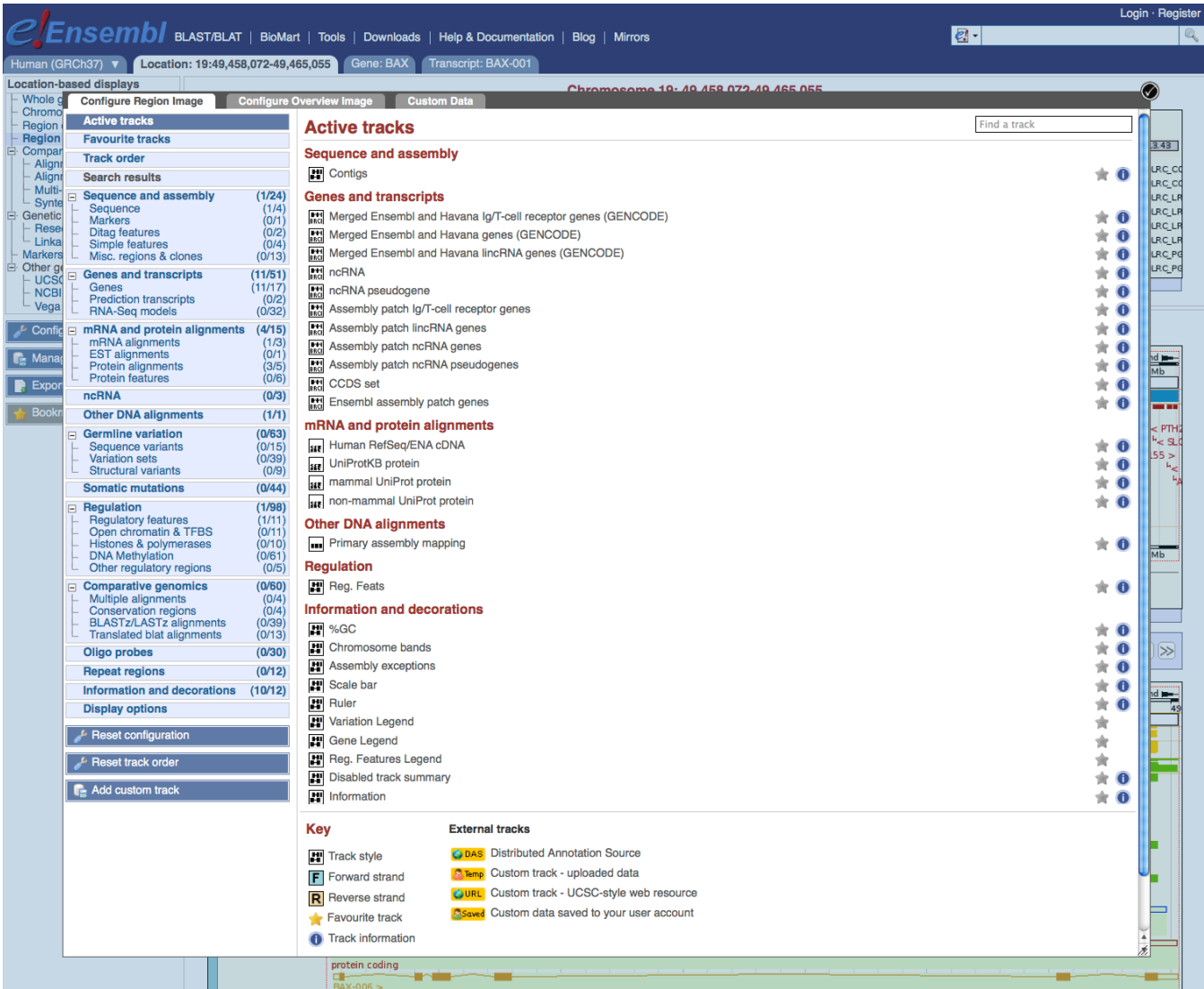


Chromosome

Region in detail

Gene in detail

To configure the location page – click ‘configure this page’ on the left hand side.



Again – a vast amount of data that can be viewed. Click on a couple of tracks to explore what the underlying information is.

When you have selected additional tracks – hit the ‘check’ mark at the top right hand corner.

Again – a vast amount of data that can be viewed. Click on a couple of tracks to explore what the underlying information is.

When you have selected additional tracks – hit the ‘check’ mark at the top right hand corner.

Worked example #2 Data download using Biomart

1. Identify all genes in a region known to be linked to a disease.
e.g. Chr1: 50 Mb – 60 Mb.

The screenshot shows the Ensembl BioMart interface. The top navigation bar includes links for BLAST/BLAT, BioMart, Tools, and More. A search bar is available for searching across all species. The main interface is divided into several sections:

- Dataset Selection:** A dropdown menu is open, showing options like "Ensembl Genes 79", "Ensembl Variation 79", "Ensembl Regulation 79", and "Vega 59". A callout box labeled "1. Choose database and dataset" points to this menu.
- Dataset List:** A list of datasets is shown, including "Homo sapiens genes (GRCh38.p2)", "Mus musculus genes (GRCm38.p3)", and "Rattus norvegicus genes (Rnor_5.0)". A callout box labeled "2. Filter: (region of interest)" points to this list.
- Filters:** A section for filtering results, including "Chromosome" (set to 1) and "Base pair" (Gene Start: 50000000, Gene End: 60000000). A callout box labeled "2. Filter: (region of interest)" points to this section.
- Count:** A button labeled "Count" is visible, with a callout box labeled "Press count to determine how many genes in the region" pointing to it.
- Results:** A summary of the selected dataset: "Dataset 231 / 64798 Genes", "Homo sapiens genes (GRCh38)", "Chromosome: 1", "Gene Start (bp): 50000000", "Gene End (bp): 60000000".

Arrows indicate the flow of the process: selecting the database and dataset, then applying filters, and finally counting the results.

3. **Attributes:** Annotations of interest Click attributes on the left hand side menu.

GENE:

Ensembl

- | | |
|--|---|
| ☒ Ensembl Gene ID | ☐ Associated Gene Source |
| ☐ Ensembl Transcript ID | ☐ Associated Transcript Name |
| ☐ Ensembl Protein ID | ☐ Associated Transcript Source |
| ☐ Ensembl Exon ID | ☐ Transcript count |
| ☐ Description | ☐ % GC content |
| ☒ Chromosome Name | ☐ Gene Biotype |
| ☐ Gene Start (bp) | ☐ Transcript Biotype |
| ☐ Gene End (bp) | ☐ Source (gene) |
| ☐ Strand | ☐ Source (transcript) |
| ☐ Band | ☐ Status (gene) |
| ☐ Transcript Start (bp) | ☐ Status (transcript) |
| ☐ Transcript End (bp) | ☐ Version (gene) |
| ☐ Transcription Start Site (TSS) | ☐ Version (transcript) |
| ☒ Associated Gene Name | |

Under the 'Gene' category, select Ensembl Gene Id, Chromosome name and associated gene name

EXTERNAL:

GO

- | | |
|---|--|
| ☐ GO Term Accession | ☐ GO Term Evidence Code |
| ☐ GO Term Name | ☐ GO domain |
| ☐ GO Term Definition | |

GOSlim GOA

- | | |
|--|---|
| ☐ GOSlim GOA Accession(s) | ☐ GOSlim GOA Description |
|--|---|

External References (max 3)

- | | |
|--|--|
| ☐ ArrayExpress | ☐ MIM Morbid Accession |
| ☐ ChEMBL ID(s) | ☐ MIM Morbid Description |
| ☐ Clone based Ensembl gene name | ☐ MIM Gene Accession |
| ☐ Clone based Ensembl transcript name | ☐ MIM Gene Description |
| ☐ Clone based VEGA gene name | ☐ miRBase Accession(s) |
| ☐ Clone based VEGA transcript name | ☐ miRBase ID(s) |
| ☐ CCDS ID | ☐ miRBase transcript name |
| ☐ Database of Aberrant 3' Splice Sites (DBASS3) IDs | ☐ Protein (Genbank) ID |
| ☐ DBASS3 Gene Name | ☐ RefSeq mRNA [e.g. NM_001195597] |
| ☐ Database of Aberrant 5' Splice Sites (DBASS5) IDs | ☐ RefSeq mRNA predicted [e.g. XM_001125684] |
| ☐ DBASS5 Gene Name | ☐ RefSeq ncRNA [e.g. NR_002834] |
| ☐ EMBL (Genbank) ID | ☐ RefSeq ncRNA predicted [e.g. XR_108264] |
| ☐ Ensembl Human Gene IDs | ☐ RefSeq Protein ID [e.g. NP_001005353] |
| ☐ Ensembl Human Transcript IDs | ☐ RefSeq Predicted Protein ID [e.g. XP_001720922] |
| ☐ Ensembl Human Translation IDs | ☐ Rfam ID |
| ☐ LRG to Ensembl link gene | ☐ Rfam transcript name |
| ☐ LRG to Ensembl link transcript | ☐ UCSC ID |
| ☒ EntrezGene ID | ☐ Unigene ID |
| ☐ EntrezGene transcript name ID | ☐ UniProt/TrEMBL Accession |
| ☐ Human Protein Atlas Antibody ID | ☐ UniProt/SwissProt ID |
| ☐ VEGA gene ID(s) (OTTG) | ☐ UniProt/SwissProt Accession |
| ☐ VEGA transcript ID(s) (OTTT) | ☐ UniProt Gene Name |
| ☐ VEGA protein ID(s) (OTTP) | ☐ Uniprot Genename Transcript Name |
| ☐ HGNC ID(s) | ☐ UniParc |
| ☒ HGNC symbol | ☐ WikiGene Name |
| ☐ HGNC transcript name | ☐ WikiGene ID |
| ☐ MEROPS ID | ☐ WikiGene Description |
| ☐ PDB ID | |

Under the External category (sub category external refs), select EntrezGene ID and HGNC symbol

Make sure the menu on the left hand side has all that you expect and then press 'results'.

New **Count** **Results**

Dataset 231 / 64798 Genes
Homo sapiens genes (GRCh38)

Filters
Chromosome: 1
Gene Start (bp): 50000000
Gene End (bp): 60000000

Attributes
Ensembl Gene ID
Chromosome Name
Associated Gene Name
EntrezGene ID
HGNC symbol

Dataset
[None Selected]

Select format (XLS) and press 'Go'

Export all results to ☒ Unique results only

Email notification to

View rows as ☐ Unique results only

Ensembl Gene ID	Chromosome Name	Associated Gene Name	EntrezGene ID	HGNC symbol
ENSG00000225030	1	RP4-784A16.3		
ENSG00000162385	1	MAGOH	4116	MAGOH
ENSG00000162385	1	MAGOH	55110	MAGOH
ENSG00000239007	1	RNU7-95P		RNU7-95P
ENSG00000252032	1	RNU6-1281P		RNU6-1281P
ENSG00000225183	1	RP4-758J24.4		
ENSG00000206627	1	RNU6-969P		RNU6-969P
ENSG00000265535	1	RN7SL475P		RN7SL475P
ENSG00000275816	1	MIR6500	102466656	MIR6500
ENSG00000264208	1	MIR4781	100616315	MIR4781

	A	B	C	D	E
1	Ensembl Gene ID	Chromosome Name	Associated Gene Name	EntrezGene ID	HGNC symbol
2	ENSG00000225030	1	RP4-784A16.3		
3	ENSG00000162385	1	MAGOH	4116	MAGOH
4	ENSG00000162385	1	MAGOH	55110	MAGOH
5	ENSG00000239007	1	RNU7-95P		RNU7-95P
6	ENSG00000252032	1	RNU6-1281P		RNU6-1281P
7	ENSG00000225183	1	RP4-758J24.4		
8	ENSG00000206627	1	RNU6-969P		RNU6-969P
9	ENSG00000265535	1	RN7SL475P		RN7SL475P
10	ENSG00000275816	1	MIR6500	102466656	MIR6500
11	ENSG00000264208	1	MIR4781	100616315	MIR4781
12	ENSG00000264081	1	AL136985.1		
13	ENSG00000279049	1	AL353898.1		
14	ENSG00000266150	1	MIR4711	100616409	MIR4711
15	ENSG00000264128	1	RN7SL713P		RN7SL713P
16	ENSG00000265538	1	MIR4421	100616189	MIR4421
17	ENSG00000264613	1	RN7SL290P		RN7SL290P
18	ENSG00000264834	1	MIR1273F	100616156	MIR1273F
19	ENSG00000230812	1	LINC01358		LINC01358
20	ENSG00000231095	1	GYG1P3		GYG1P3
21	ENSG00000207194	1	RNU6-1026P		RNU6-1026P
22	ENSG00000265815	1	MIR1273G	100616145	MIR1273G
23	ENSG00000266188	1	MIR5095	100616458	MIR5095
24	ENSG00000231090	1	RP11-101C11.1		
25	ENSG00000272051	1	Y_RNA		
26	ENSG00000223307	1	AC119674.1		
27	ENSG00000199831	1	RN7SKP291		RN7SKP291
28	ENSG00000234578	1	RP4-784A16.1		
29	ENSG00000279054	1	AL445183.1		
30	ENSG00000200839	1	RNA5SP48		RNA5SP48
31	ENSG00000221591	1	AL161740.1		
32	ENSG00000157193	1	LRP8	7804	LRP8
33	ENSG00000238773	1	AL137855.1		
34	ENSG00000200262	1	Y_RNA		
35	ENSG00000265404	1	AC099796.1		
36	ENSG00000201003	1	SNORA58		

Open up file in Excel

There is a lot more you can do with Biomart.

As a general rule: if you can view data in Ensembl – it can be exported!

The Genome Access Course

Comparative Genome Analysis



The Genome Access Course

Comparative Genome Analysis

April 2015

The Genome Access Course

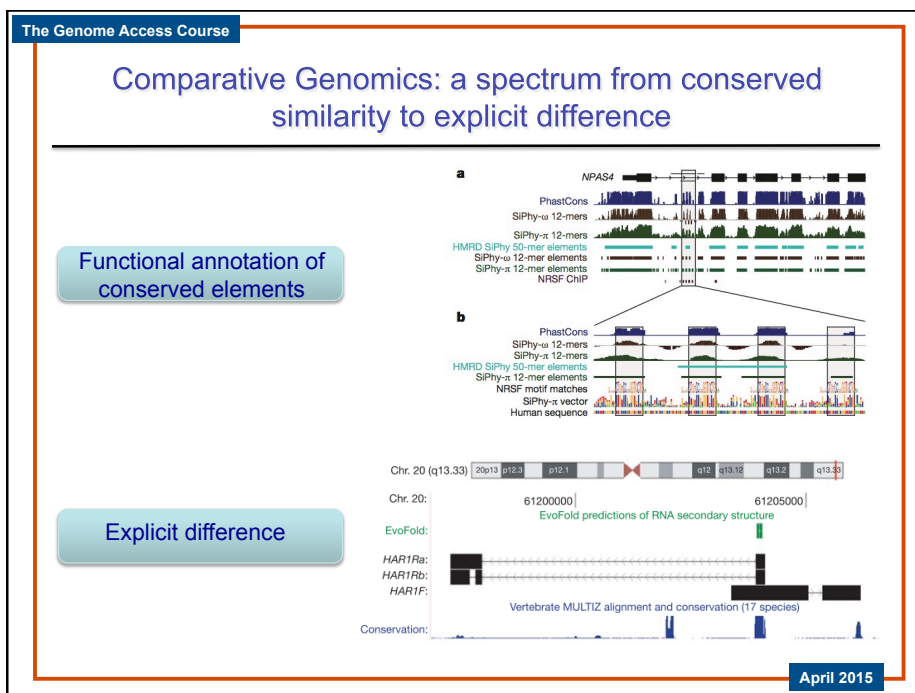
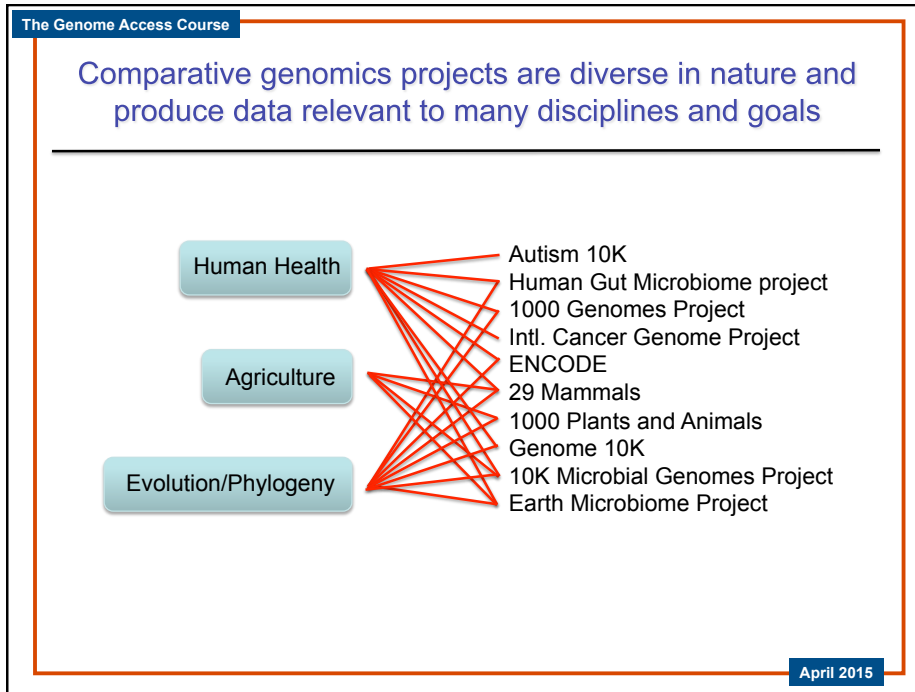
Comparative Genome Analysis

What is it?

Selected elements across taxa are informative for:

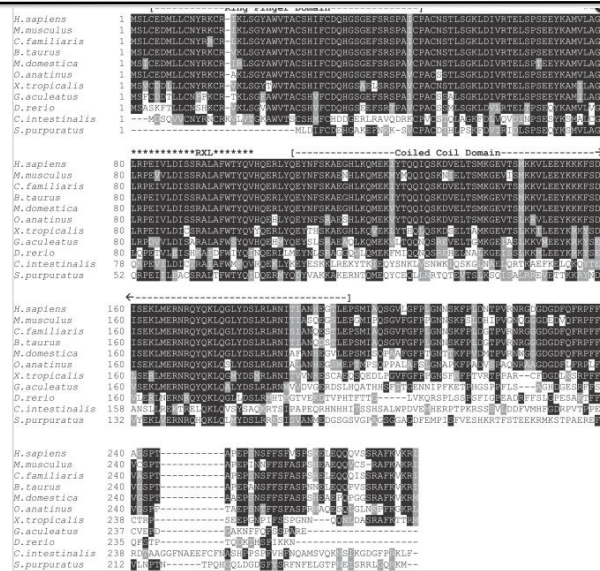
- Gene identification in novel genomes
- Genome organization studies
- Homology, paralogy, and orthology
- Functional element identification (coding and non-coding; within and among species)*
- Species specific characteristics*
- Comparative epigenomics
- Molecular evolution studies
- And others...

April 2015



The Genome Access Course

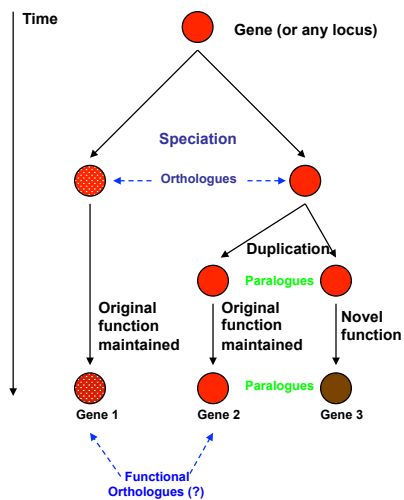
Important function can be revealed by both similarity and difference



April 2015

The Genome Access Course

Identifying homologs, orthologs, paralogs: the relationship of derived and ancestral sequences



Homologs - Genes derived from common ancestral gene

Orthologs – Genes in *different species* that are derived from the same gene in last common ancestor

Paralogs – Gene families that have diverged *within a single species*, often by duplication

April 2015

The Genome Access Course

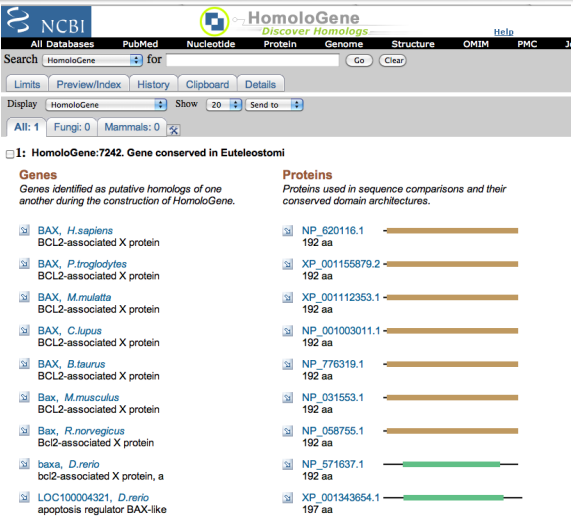
The species selected will inform and influence the genome comparison results: shallow or deep comparison?

				
Human vs.	Chimpanzee	Mouse	Opossum	Pufferfish
Size (Gbp)	3.0	2.5	4.2	0.4
Time since divergence	~6 MYA	~ 90 MYA	~150 MYA	~450 MYA
Sequence conservation (in coding regions)	>99%	~80%	~70-75%	~65%
Aids identification of...	Recently changed sequences and genomic rearrangements	Both coding and non-coding sequences	Both coding and non-coding sequences	Primarily coding sequences
Background noise	High	Moderate	Low	Lower

April 2015

The Genome Access Course

Among several options, genome browsers provide very good initial ortholog prediction - NCBI



The screenshot shows the NCBI HomoloGene search results for the query 'BCL2-associated X protein'. The interface includes a search bar, navigation tabs (All Databases, PubMed, Nucleotide, Protein, Genomes, Structure, OMIM, PMC), and a results section. The results are organized into two columns: 'Genes' and 'Proteins'. The 'Genes' column lists various BCL2-associated X proteins from different species, including *H. sapiens*, *P. troglodytes*, *M. mulatta*, *C. lupus*, *B. taurus*, *M. musculus*, *R. norvegicus*, *D. rerio*, and *D. rerio*. The 'Proteins' column lists the corresponding protein sequences, such as NP_620116.1, XP_001155879.2, XP_001112353.1, NP_001003011.1, NP_776319.1, NP_031553.1, NP_058755.1, NP_571637.1, and XP_001343654.1. The results are filtered by 'All: 1', 'Fungi: 0', and 'Mammals: 0'.

April 2015

The Genome Access Course

Ensembl uses phylogeny (via the GeneTree view) to compute paralogs and orthologs

Export Image size

Species	Gene ID	Score
OR6 M1 Pan troglodytes		
OR6 M1 Homo sapiens		
OR6 M1 Macaca mulatta		
OR6 M1 Otdemur gametti		
489350 Canis familiaris		
OR6 M1 Dasyus novemcinctus		
OR6 M1 Erinaceus europaeus		
489349 Canis familiaris		
489348 Canis familiaris		
OR6 M1 Felis catus		
489347 Canis familiaris		
OR6 M1 Monodelphis domestica		

April 2015

The Genome Access Course

Ensembl uses phylogeny (via the GeneTree view) to compute paralogs and orthologs

Orthologues

View sequence alignments of all orthologues

Summary of orthologues of this gene

Click on 'Show' to display the orthologues for one or more groups, or click on 'Configure this page' to choose a custom list of species

Species set	Show details	1:1	1:many	many:many	No orthologues
Primates	☒	7	0	0	2
Humans and other primates	☒	7	0	0	2
Rodents	☐	8	0	0	0
Rodents, rabbits and related species	☐	8	0	0	0
Laursiatheria	☐	10	0	0	3
Carnivores, ungulates and insectivores	☐	10	0	0	3
Placental Mammals	☐	27	0	0	8
All placental mammals	☐	27	0	0	8
Sauropsids	☐	5	0	0	0
Birds and Reptiles	☐	5	0	0	0
Fish	☐	3	5	0	1
Ray-finned fishes	☐	3	5	0	1
All	☐	40	11	0	14
All species, including invertebrates	☐	40	11	0	14

Selected orthologues

Species	Type	dN/dS	Ensembl identifier & gene name	Compare	Location	Target %id	Query %id
Bushbaby (<i>Otilomys garnettii</i>)	1-to-1	0.40145	ENSOGAG000000000338 VEGFA vascular endothelial growth factor A (Source:HGNC Symbol;Acc:10860)	- Region - Comparison - Alignment (protein) - Alignment (cDNA) - Gene Tree (image)	GL873550.1:14291591-14304782.1	87	49
Gibbon (<i>Nomascus leucogenys</i>)	1-to-1	0.00000	ENSNGAG0000000003829 VEGFA Uncharacterized protein (Source:UniProtKB/TrEMBL;acc:Q1G2N7)	- Region - Comparison - Alignment (protein) - Alignment (cDNA) - Gene Tree (image)	GL397304.1:16013522-16027357.1	100	45

April 2015

The Genome Access Course

Worked example #1

April 2015

The Genome Access Course

Important function can be revealed by both similarity and difference

Using example numbers

Sequence alignment showing conserved regions (indicated by asterisks) and a Coiled Coil Domain (indicated by a bracket). The alignment includes sequences from various species, such as H. sapiens, M. musculus, C. familiaris, B. taurus, W. domesticus, O. anatinus, X. tropicalis, G. sculeatus, D. rerio, C. intestinalis, and S. purpuratus.

April 2015

For a more customized analysis of coding sequence one may construct a multiple sequence alignment and phylogenetic tree

A. Obtain sequences from all species to compare

Genome browsers, **Uniprot**, PCR amplify and sequence etc.

B. Align all sequences

MUSCLE, **ClustalW**, DIALIGN, T-Coffee, etc.

C. Use different methods to reconstruct the evolutionary history of the sequences

PAUP (Parsimony), Likelihood (Garli), Bayesian (Mr. Bayes),

Distance (**Neighbor-joining**)

** NOTE: each of these methods have pros and cons and it is often necessary to try several.

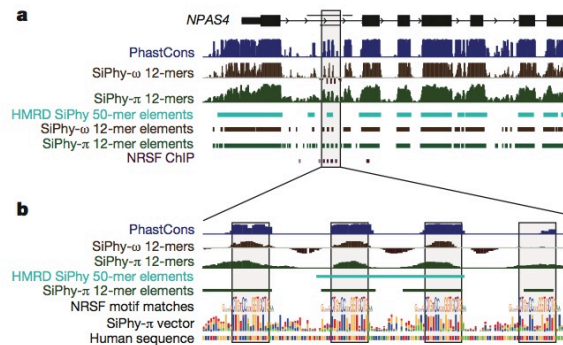
April 2015

Worked example #2

April 2015

The Genome Access Course

Beyond analysis of a single gene, functionally conserved units can be identified at the sequence level



Layering subsequent genome annotations will increase the resolution

- ESTs
- ChIP Seq data
- cDNA
- Etc...

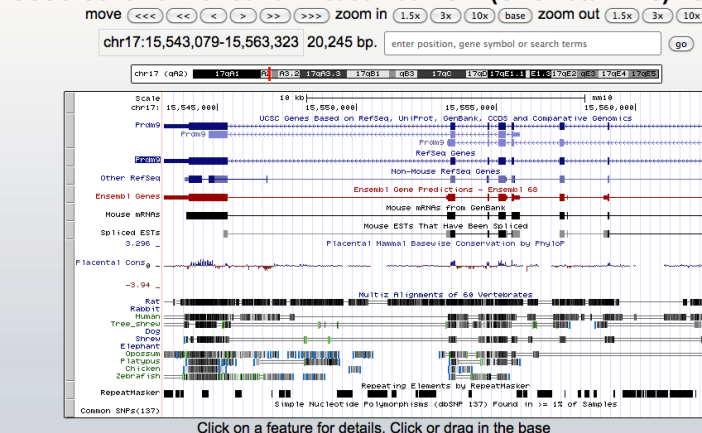
Conserved sequence can also be identified in non-coding regions as well

April 2015

The Genome Access Course

Genome browsers such as UCSC provide a broad range of comparative genomic tools (general browser and custom)

UCSC Genome Browser on Mouse Dec. 2011 (GRCm38/mm10) Assembly



April 2015

The Genome Access Course

Genome browsers such as UCSC provide a broad range of comparative genomic tools (general browser and custom)

Conservation Track Settings

Vertebrate Multiz Alignment & PhastCons Conservation (28 Species)

Display mode:

Conservation: ☒ Placental Mammal ☐ Vertebrate

Pairwise alignments:

Primate

☒ Chimp ☒ Rhesus ☒ Bushbaby

Placental Mammal

☐ TreeShrew ☒ Mouse ☐ Rat ☐ GuineaPig ☐ Rabbit

☐ Shrew ☐ Hedgehog ☒ Dog ☒ Cat ☒ Horse

☐ Cow ☒ Armadillo ☐ Elephant ☐ Tenrec

Vertebrate

☒ Opossum ☒ Platypus ☐ Lizard ☐ Chicken ☐ X. tropicalis

☐ Zebrafish ☐ Tetraodon ☐ Fugu ☐ Stickleback ☐ Medaka

Multiple alignment base-level:

☒ Display bases identical to reference as dots

☒ Display chains between alignments

Codon Translation:

Default species to establish reading frame:

☐ No codon translation

☒ Use default species reading frames for translation

☐ Use reading frames for species if available, otherwise no translation

☐ Use reading frames for species if available, otherwise use default species

April 2015

The Genome Access Course

Worked example #3

April 2015

WE1

A. Identify orthologues of human BAX gene in Ensembl

Gene: BAX ENSG00000087088

Description BCL2-associated X protein [Source:HGNC Symbol;Acc:HGNC:959]

Synonyms BCL2L4

Location [Chromosome 19: 48,954,815-48,961,798](#) forward strand.

INSDC coordinates chromosome:GRCh38:CM000681.2:48954815:48961798:1

Transcripts

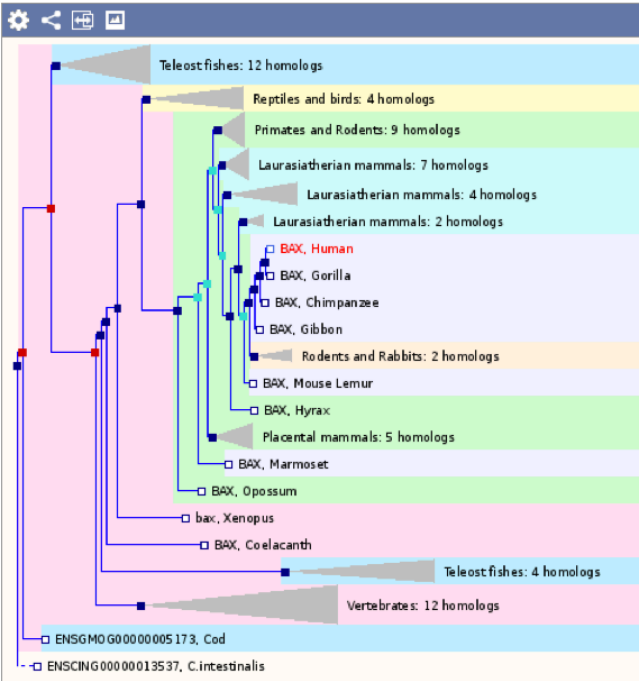
This gene has 13 transcripts (splice variants) [Show transcript table](#)

Human BAX gene

Gene-based displays

- Summary
- Splice variants (13)
- Transcript comparison
- Supporting evidence
- Alt. alleles
- Sequence
 - Secondary Structure
- External references
- Regulation
- Comparative Genomics
 - Genomic alignments
 - Gene tree (image)
 - Gene tree (text)
 - Gene tree (alignment)
 - Gene gain/loss tree
 - Orthologues (56)
 - Paralogues
 - Protein families (2)
- Phenotype
- Genetic Variation
 - Variation table
 - Variation image
 - Structural variation
- External data
 - Personal annotation
- ID History
 - Gene history

View Gene tree (image)



LEGEND						
Branch Length	Nodes	Genes	Collapsed nodes	Collapsed Alignments	Expanded Alignments	
— x1 branch length	□ gene node	Gene ID	■ collapsed sub-tree	□ 0 - 33% aligned seq	□ gap	
- - - x10 branch length	■ speciation node	gene of interest	■ collapsed (this gene)	■ 33 - 66% aligned seq	■ aligned seq	
- . - x100 branch length	■ duplication node	Gene ID within-sp. paralog	■ collapsed (paralog)	■ 66 - 100% aligned seq		
	■ ambiguous node		■ (x10 branch length)			
	■ gene split event		■ (x100 branch length)			

Gene: **BAX** ENSG00000087088

Description

Synonyms

Location

INSDC coordinates

Transcripts

BCL2-associated X protein [Source:HGNC Symbol;Acc:HGNC:959]

BCL2L4

[Chromosome 19: 48,954,815-48,961,798](#) forward strand.

chromosome:GRCh38:CM000681.2:48954815:48961798:1

This gene has 13 transcripts (splice variants) [Show transcript table](#)

Human BAX gene

Gene-based displays

Summary

Splice variants (13)

Transcript comparison

Supporting evidence

Alt. alleles

Sequence

Secondary Structure

External references

Regulation

Comparative Genomics

Genomic alignments

Gene tree (image)

Gene tree (text)

Gene gain/loss tree

Orthologues (56)

Paralogues

Protein families (2)

Phenotype

Genetic Variation

Variation in structure

External data

Personal annotation

ID History

Gene history

View Orthologues

Orthologues ⓘ

[View sequence alignments of all orthologues.](#)

Summary of orthologues of this gene

Click on 'Show' to display the orthologues for one or more groups, or click on 'Configure this page' to choose a custom list of species

Species set	Show details	1:1	1:many	many:many	No orthologues
Primates Humans and other primates	☐	11	0	0	0
Rodents Rodents, rabbits and related species	☒	7	0	0	0
Laurasiatheria Carnivores, ungulates and insectivores	☐	12	1	0	1

Selected orthologues

Show **All** entries

Show/hide columns

Filter

Species	Type	dN/dS	Ensembl identifier & gene name	Compare	Location	Target %id	Query %id
Guinea Pig (<i>Cavia porcellus</i>)	1-to-1	0.18974	ENSCPOG000000011196 BAX BCL2-associated X protein [Source:HGNC Symbol;Acc:HGNC:959]	• Region Comparison • Alignment (protein) • Alignment (cDNA) • Gene Tree (image)	scaffold_70:8361550-8365985:-1	78	68
Kangaroo rat (<i>Dipodomys ordii</i>)	1-to-1	n/a	ENSDORG000000015554 Bax BCL2-associated X protein [Source:MGI Symbol;Acc:MGI:99702]	• Region Comparison • Alignment (protein) • Alignment (cDNA) • Gene Tree (image)	scaffold_3122:67780-76		
Mouse (<i>Mus musculus</i>)	1-to-1	0.18690	ENSMUSG000000003873 Bax BCL2-associated X protein [Source:MGI Symbol;Acc:MGI:99702]	• Region Comparison • Alignment (protein) • Alignment (cDNA) • Gene Tree (image)	7:45461695-45466898:-1	78	68
Pika (<i>Ochotona princeps</i>)	1-to-1	n/a	ENSOPRG000000002468 BAX BCL2-associated X protein [Source:HGNC Symbol;Acc:HGNC:959]	• Region Comparison • Alignment (protein) • Alignment (cDNA) • Gene Tree (image)	GeneScaffold_992:74359-77560:1	76	68

Download what you see or whole table

2

WE2

B. Download orthologues for a list of genes

New Count Results URL XML Perl Help

Dataset
Homo sapiens genes (GRCh38)

Filters
Chromosome: 1
Gene Start (bp): 50000000
Gene End (bp): 60000000

Attributes
Ensembl Gene ID
Ensembl Transcript ID

Dataset
[None Selected]

Please restrict your query using criteria below

REGION:

☒ Chromosome

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20

☒ Base pair
Gene Start (bp) 50000000
Gene End (bp) 60000000

Select region to filter:
Chr1
50000000-60000000
(can also be a list of genes)

☐ Features ☐ Variation
☐ Structures ☐ Sequences
☒ Homologs

GENE:

Ensembl

☒ Ensembl Gene ID
☐ Ensembl Transcript ID
☐ Ensembl Protein ID
☐ Chromosome Name
☐ Gene Start (bp)
☐ Gene End (bp)
☐ Strand

☐ Band
☒ Associated Gene Name
☐ Associated Gene Source
☐ Transcript count
☐ % GC content
☐ Description

Select **Homologs**
from attributes menu

Choose attributes:
For human genes

Caenorhabditis elegans Orthologs

☐ Caenorhabditis elegans Chromosome Start (bp)
☒ Caenorhabditis elegans Ensembl Gene ID
☐ Canonical Protein or Transcript ID
☐ Caenorhabditis elegans Ensembl Protein ID
☒ Caenorhabditis elegans Chromosome Name
☐ Caenorhabditis elegans Chromosome End (bp)
☐ Homology Type

☐ Ancestor
☒ Orthology confidence [0 low, 1 high]
☒ % Identity with respect to query gene
☐ % Identity with respect to Caenorhabditis elegans gene
☐ dN
☐ dS

Choose orthologues
to download (e.g. C.
elegans and mouse).

Mouse Orthologs

☒ Mouse Ensembl Gene ID
☐ Canonical Protein or Transcript ID
☐ Mouse Ensembl Protein ID
☒ Mouse Chromosome Name
☐ Mouse Chromosome Start (bp)
☐ Mouse Chromosome End (bp)
☐ Homology Type

☐ Ancestor
☒ Orthology confidence [0 low, 1 high]
☒ % Identity with respect to query gene
☐ % Identity with respect to Mouse gene
☐ dN
☐ dS

WE1

Export all results to ☒ Unique results only

Email notification to

View rows as ☐ Unique results only

Ensembl Gene ID	Associated Gene Name	Caenorhabditis elegans Ensembl Gene ID	Caenorhabditis elegans Chromosome Name	% Identity with respect to query gene	Orthology confidence [0 low, 1 high]	Mouse Ensembl Gene ID	Mouse Chromosome Name
ENSG00000225030	RP4-784A16.3						
ENSG00000162385	MAGO	WBGene00003123	I	79	1	ENSMUSG00000028803	4
ENSG00000239007	RNU7-95P						
ENSG00000252032	RNU6-1281P						
ENSG00000225183	RP4-758J24.4						
ENSG00000206627	RNU6-969P						
ENSG00000285535	RN7SL475P						
ENSG00000275816	MIR6500						
ENSG00000264208	MIR4781						
ENSG00000264081	AL136985.1						

Export results:
Unique results in XLS
file

Ensembl Gene ID	Associated Gene Name	Caenorhabditis elegans Ensembl Gene ID	Caenorhabditis elegans Chromosome Name	% Identity with respect to query gene	Orthology confidence [0 low, 1 high]	Mouse Ensembl Gene ID	Mouse Chromosome Name	Orthology confidence [0 low, 1 high]
ENSG00000225030	RP4-784A16.3							
ENSG00000162385	MAGO	WBGene00003123	I	79	1	ENSMUSG00000028803	4	
ENSG00000239007	RNU7-95P							
ENSG00000252032	RNU6-1281P							
ENSG00000225183	RP4-758J24.4							
ENSG00000206627	RNU6-969P							
ENSG00000285535	RN7SL475P							
ENSG00000275816	MIR6500							
ENSG00000264208	MIR4781							
ENSG00000264081	AL136985.1							
ENSG00000298495	AL353898.1							
ENSG00000288150	MIR4711							
ENSG00000284149	RN7SL473P							
ENSG00000282538	MIR4421							
ENSG00000284813	RN7SL470P							
ENSG00000284834	MIR1273P							
ENSG00000282812	LINC01358							
ENSG00000210395	D10103							
ENSG00000207194	RNU6-1028P							
ENSG00000285615	MIR1273D							
ENSG00000288188	MIR6505							
ENSG00000231090	RP11-101C11.1							
ENSG00000272251	Y RNA							
ENSG00000223307	AC119874.1							
ENSG00000198631	RNU5029P							
ENSG00000284578	RP4-784A16.1							
ENSG00000272504	AL445183.1							
ENSG00000228268	RNASSP48							
ENSG00000221581	AL181740.1							
ENSG00000157193	LDBP							
ENSG00000282773	AL137855.1					ENSMUSG00000028813	4	
ENSG00000202265	Y RNA							
ENSG00000284504	AC109718.1							
ENSG00000210301	SCN2A58					ENSMUSG00000028438	12	1
ENSG00000283892	MIR4422							
ENSG00000241748	RN7SL788P							
ENSG00000238640	RN7SL62P							
ENSG00000282585	RNU6-877P							
ENSG00000281156	RNU6-830P							
ENSG00000286993	RP4-657D16.3							
ENSG00000232407	RP11-587G23.3							
ENSG00000271723	MIR407-TTCA	WBGene00018418	I	14	1	ENSMUSG000000047562	4	1
ENSG00000242845	SLC25A9P3							
ENSG00000284585	RP11-133G22.1							
ENSG00000162457	PPAP2B	WBGene00020895	II	25	1	ENSMUSG00000028817	4	1
ENSG00000229116	HKSC1P48							
ENSG00000226871	PCP-1P3							
ENSG00000117559	QSBP19	WBGene00016331	I	40	1	ENSMUSG00000028559	4	1
ENSG00000224595	RP11-377G22.3							
ENSG00000238778	RPL41P23							
ENSG00000225475	RP11-377G22.2							
ENSG00000226913	RP11-378J3.1							
ENSG00000243728	TTCA	WBGene00015916	II	31		ENSMUSG000000028413	4	
ENSG00000118162	GPX2					ENSMUSG00000028597	4	
ENSG00000228638	RP4-784A16.2							
ENSG00000220754	RP5-1024G8.5							
ENSG00000207172	RP5-1024G8.8							
ENSG00000184313	MIR407	WBGene00018418	I	14	1	ENSMUSG000000047562	4	1
ENSG00000185051	MR5M1					ENSMUSG00000056297	4	1
ENSG00000233971	RP11-12C17.2							
ENSG00000143021	TME8B1							
ENSG00000202143	TLBP1D							
ENSG00000238360	RP11-334A14.2							
ENSG00000232116	RP11-41C16.1							
ENSG00000226005	C1orf185					ENSMUSG000000360491	4	1
ENSG00000118171	SC2P	WBGene00013284	II	41		ENSMUSG00000028853	4	1
ENSG00000232738	DAR1-AS1							
ENSG00000058795	YIPF1	WBGene00021294	II	34	1	ENSMUSG000000057275	4	
ENSG00000252018	RNU6-30P							
ENSG00000182395	AC0111					ENSMUSG00000034893	4	
ENSG00000213478	C11orf2							
ENSG00000217359	RP5-102G10.1							
ENSG0000021884	SRB							
ENSG00000233356	RP11-298A18.3	WBGene00017326						1
ENSG00000144729	CART3							
ENSG00000258885	ENSG00000258885							

Excel file showing all known
orthologues in worm and mouse,
confidence of prediction and % ID

WE2

C. Multiple sequence alignment and phylogenetic trees

UniProtKB

BLAST Align Upload Lists

Results

Filter by 1 to 2

Reviewed (337)
Swiss-Prot

Unreviewed (30,119)
TrEMBL

Popular organisms
Human (220)
Mouse (85)
Rat (41)
Bovine (36)
Zebrafish (34)

Other organisms

Search terms
Filter "bax" as:
author (71)
gene name (385)
gene ontology (47)
protein name (4,213)
strain (1)
taxonomy (1)

View by
Taxonomy
Keywords
Gene Ontology
Enzyme class
Pathway

UniRef
Your results in sequence clusters

Entry	Entry name	Protein names	Gene names
10 result(s) selected. (Clear selection)			
☒ Q07812	BAX_HUMAN	Apoptosis regulator BAX	BAX, BCL2L4
☒ Q63690	BAX_RAT	Apoptosis regulator BAX	Bax
☒ Q07813	BAX_MOUSE	Apoptosis regulator BAX	Bax
☒ P27297	BAX_ECOLI	Protein bax	bax, b3570, JW5653
☒ O02703	BAX_BOVIN	Apoptosis regulator BAX	BAX
☐ P55061	BI1_HUMAN	Bax inhibitor 1	TMBIM6, BI1, TEGT
☐ H0YA56	H0YA56_HUMAN	Apoptosis regulator BAX	BAX
☐ Q9D2C7	BI1_MOUSE	Bax inhibitor 1	Tmbim6, Tegt
☒ Q98U13	Q98U13_XENLA	BCL2-associated X protein	bax, bax-A
☒ Q9I9N4	Q9I9N4_DANRE	Bax	baxa, bax
☐ G3V8T9	G3V8T9_RAT	Apoptosis regulator BAX	Bax, rCG_54570
☒ F6ZV96	F6ZV96_MACMU	Apoptosis regulator BAX isoform alp...	BAX
☐ Q9JKL3	Q9JKL3_RAT	Bax protein splice variant k	Bax, Bax_v1, rCG_54570

Search BAX gene in Uniprot.

Select BAX in a variety of different organisms and

Download selections in FASTA

WE2

```
>sp|Q07812|BAX_HUMAN Apoptosis regulator BAX OS=Homo sapiens GN=BAX PE=1 SV=1
MDGSGEQPRGGGPTSSEQIMKTGALLQGFIQDRAGRMGGEAPELALDPVPQDASTKKLS
ECLKRIGDELDSDNMELQRMIAAVDTSPPREVFFRVAADMFSNGFNWGRVVALFYFASKL
VLKALCTKVPELIRTIMGWTLDFLRERLLGWIQDQGGWDGLLSYFGTPTWQTVTIFVAGV
LTASLTIWKKMG
>sp|Q63690|BAX_RAT Apoptosis regulator BAX OS=Rattus norvegicus GN=Bax PE=1
MDGSGEQLGSGGPTSSEQIMKTGAFLQGFQDRAGRMAGETPELTLEQPPQDASTKKLS
ECLRRIGDELDSDNMELQRMIAAVDTSPPREVFFRVAADMFSNGFNWGRVVALFYFASKL
VLKALCTKVPELIRTIMGWTLDFLRERLLVWQIQDQGGWDGLLSYFGTPTWQTVTIFVAGV
LTASLTIWKKMG
>sp|O02703|BAX_BOVIN Apoptosis regulator BAX OS=Bos taurus GN=BAX PE=2 SV=1
MDGSGEQPRGGGPTSSEQIMKTGALLQGFIQDRAGRMGGETPELGLQVPQDASTKKLS
ECLKRIGDELDSDNMELQRMIAAVDTSPPREVFFRVAADMFSNGFNWGRVVALFYFASKL
VLKALCTKVPELIRTIMGWTLDFLRERLLGWIQDQGGWDGLLSYFGTPTWQTVTIFVAGV
LTASLTIWKKMG
>sp|P27297|BAX_ECOLI Protein bax OS=Escherichia coli (strain K12) GN=bax PE=
MILTPIRRYGAMILMLLTLVFSSEVLAKHTTTASQKSHLTAKSNKQVSSKQEYSRNSAK
SSSLPDLRKYPSTPRKKAFLRTVMPYITSQNAAITAERNWLISKQYQGWSPAERARLK
DIAKRYKVKWSGNTRKIPWNTLLERVDIIPSTMVATMAAASGWSKSLARNNNLFGMK
CMKGRCTNAPGKVKGYSQFSSVKESVSAYVTNLNTHPAYSSFRKSRAQLRKADQEVTTATA
MIHKLKGYSTKGKSYNNYLFAMYQDNQRLIAAHM
>sp|Q07813|BAX_MOUSE Apoptosis regulator BAX OS=Mus musculus GN=Bax PE=1 SV=1
MDGSGEQLGSGGPTSSEQIMKTGAFLQGFQDRAGRMAGETPELTLEQPPQDASTKKLS
ECLRRIGDELDSDNMELQRMIAAVDTSPPREVFFRVAADMFSNGFNWGRVVALFYFASKL
VLKALCTKVPELIRTIMGWTLDFLRERLLVWQIQDQGGWDGLLSYFGTPTWQTVTIFVAGV
LTASLTIWKKMG
>tr|F6ZV96|F6ZV96_MACMU Apoptosis regulator BAX isoform alpha OS=Macaca mulatta GN=BAX PE=2 SV=1
MDGSGEQPRGGGPTSSEQIMKTGALLQGFIQDRAGRMGGETPELALDPVPQDASTKKLS
ECLKRIGDELDSDNMELQRMIAAVDTSPPREVFFRVAADMFSNGFNWGRVVALFYFASKL
VLKALCTKVPELIRTIMGWTLDFLRERLLGWIQDQGGWDGLLSYFGTPTWQTVTILVAGV
LTASLTIWKKMG
>tr|Q9I9N4|Q9I9N4_DANRE Bax OS=Danio rerio GN=baxa PE=2 SV=1
MAAPSGGGDTGSGNDQILDGAALLNNFVYERVRHGRDRDAEVTRSQLGGLVCLDPHSHKR
LAQCLQQIGDELGNALQSMNLNNSLQPTQDVFIRVAREIFSNGKFNWGRVVALFYFAC
RLVKAISTRVDPDIIRTIISWTMSYIQEHVINWIREQGGWDGIRSYFGTPTWQTVGVFLA
GVITTALVIRKM
>tr|Q9GMG7|Q9GMG7_SHEEP Bcl2-associated protein Bax (Fragment) OS=Ovis aries GN=Bax PE=2 SV=1
GFIQDRDRMGGETPELGLQVPQDASTKKLSECLKRIGDELDSDNMELQRMIAAVDTSPP
REVFFRVAADMFSNGFNWGRVVALFYFASKLVLKALCTKVPELIRTIMGWTLDFLRERL
LGWQIQDQGGWDGLLSYFGTPTWQTVTIFV
>tr|Q70AD3|Q70AD3_PIG Bax-alpha protein (Fragment) OS=Sus scrofa GN=BAX-alpha PE=2 SV=1
FIQDRAGRMGGETPELGLQVPQDASTKKLSECLKRIGDELDSDNMELQRMIAAVDTSPP
EVE
GWI
>tr
MAT
RLQ
VFE
WIF
```

FASTA file for
protein
sequences for
selected
organisms.

Copy the results
to paste into
CLUSTALW

ClustalW2

Input form Web services Help & Documentation

Share Feedback

Tools > Multiple Sequence Alignment > ClustalW2

Multiple Sequence Alignment

ClustalW2 is a general purpose DNA or protein multiple sequence alignment program for three or more sequences. If you only have two sequences please instead use our [pairwise sequence alignment tools](#).

Note: **ClustalW2 is no longer being maintained.** Please consider using the new version instead.

STEP 1 - Enter your input sequences

Enter or paste a set of sequences in any supported format:

```
>sp|Q07812|BAX_HUMAN Apoptosis regulator BAX OS=Homo sapiens GN=BAX PE=1 SV=1
MDGSGEQPRGGGPTSSEQIMKTGALLQGFIQDRAGRMGGEAPELALDPVPQDASTKKLS
ECLKRIGDELDSDNMELQRMIAAVDTSPPREVFFRVAADMFSNGFNWGRVVALFYFASKL
VLKALCTKVPELIRTIMGWTLDFLRERLLGWIQDQGGWDGLLSYFGTPTWQTVTIFVAGV
LTASLTIWKKMG
>sp|Q63690|BAX_RAT Apoptosis regulator BAX OS=Rattus norvegicus GN=Bax PE=1 SV=1
MDGSGEQLGSGGPTSSEQIMKTGAFLQGFQDRAGRMAGETPELTLEQPPQDASTKKLS
ECLRRIGDELDSDNMELQRMIAAVDTSPPREVFFRVAADMFSNGFNWGRVVALFYFASKL
VLKALCTKVPELIRTIMGWTLDFLRERLLVWQIQDQGGWDGLLSYFGTPTWQTVTIFVAGV
LTASLTIWKKMG
```

Or, upload a file: No file selected.

STEP 4 - Submit your job

☐ Be notified by email (Tick this box if you want to be notified by email when the results are available)

Paste into
ClustalW at EBI,
scroll down and
hit **Submit**

WE2

ClustalW2

Input form | Web services | Help & Documentation

Tools > Multiple Sequence Alignment > ClustalW2

Results for job clustalw2-l20141110-155348-0272-65017105-es

Alignments | Result Summary | Guide Tree | Phylogenetic Tree | Submission Details

Download Alignment File | Show Colors | Send to ClustalW2_Phylogeny

Results page:
Show colors

CLUSTAL 2.1 multiple sequence alignment

```
sp|Q07812|BAX_HUMAN      -----MDGSGEQPRG 10
tr|F6ZV96|F6ZV96_MACMU  -----MDGSGEQPRG 10
sp|O02703|BAX_BOVIN      -----MDGSGEQPRG 10
tr|Q9GMG7|Q9GMG7_SHEEP   -----
sp|Q63690|BAX_RAT        -----
sp|Q07813|BAX_MOUSE      -----
tr|Q70AD3|Q70AD3_PIG     -----
tr|Q98U13|Q98U13_XENLA   -----
tr|Q9I9N4|Q9I9N4_DANRE   -----
sp|P27297|BAX_ECOLI      -----
```

Download Alignment File | Hide Colors | Send to ClustalW2_Phylogeny

Send to
ClustalW2_Phylogeny

CLUSTAL 2.1 multiple sequence alignment

```
sp|Q07812|BAX_HUMAN      -----MDGSGEQPRG 10
tr|F6ZV96|F6ZV96_MACMU  -----MDGSGEQPRG 10
sp|O02703|BAX_BOVIN      -----
tr|Q9GMG7|Q9GMG7_SHEEP   -----
sp|Q63690|BAX_RAT        -----
sp|Q07813|BAX_MOUSE      -----
tr|Q70AD3|Q70AD3_PIG     -----
tr|Q98U13|Q98U13_XENLA   -----
tr|Q9I9N4|Q9I9N4_DANRE   -----
sp|P27297|BAX_ECOLI      -----
```

```
sp|Q07812|BAX_HUMAN      -----GGPTSSEQIMKTGALLQGFIQDRAGR---MGGEAPELAL 47
tr|F6ZV96|F6ZV96_MACMU  -----GGPTSSEQIMKTGALL
sp|O02703|BAX_BOVIN      -----GGPTSSEQIMKTGALL
tr|Q9GMG7|Q9GMG7_SHEEP   -----GGPTSSEQIMKTGAFI
sp|Q63690|BAX_RAT        -----GGPTSSEQIMKTGAFI
sp|Q07813|BAX_MOUSE      -----GGPTSSEQIMKTGAFI
```

Automatically enters
the alignment into
phylogeny

Clustering method:
Neighbour-joining

Submit

Phylogeny

Commonly used phylogenetic tree generation methods provided by the ClustalW2 program.

STEP 1 - Enter your multiple sequence alignment

Enter or paste a multiple sequence alignment in any supported format:

clustalw2-l20141110-155348-0272-65017105-es

Or, upload a file: No file selected.

STEP 2 - Set your Phylogeny options

TREE FORMAT	DISTANCE CORRECTION	EXCLUDE GAPS	CLUSTERING METHOD	P.I.M.
Default	off	off	Neighbour-joining	off

STEP 3 - Submit your job

☐ Be notified by email (Tick this box if you want to be notified by email when the results are available)

WE2

Phylogenetic Tree

[View Phylogenetic Tree File](#)

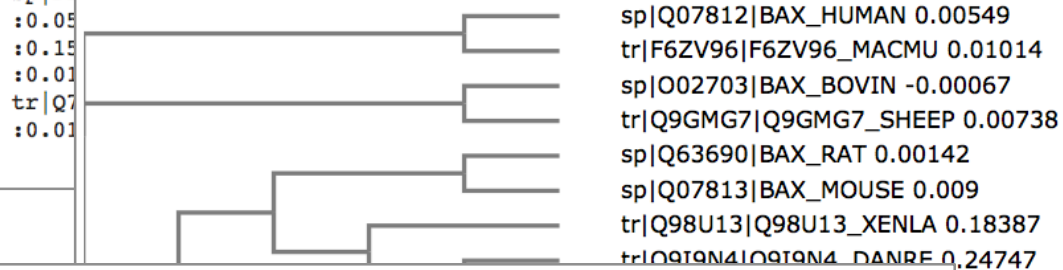
```
(
(
sp|Q07812|BAX_HUMAN:0.00549,
tr|F6ZV96|F6ZV96_MACMU:0.01014)
:0.02086,
(
sp|O02703|BAX_BOVIN:-0.00067,
tr|Q9GMG7|Q9GMG7_SHEEP:0.00738)
:0.00411,
(
(
(
sp|Q63690|BAX_RAT:0.00142,
sp|Q07813|BAX_MOUSE:0.00900)
:0.02668,
```

Phylogenetic tree (generic format)

Cladogram
(unit spacing)

Phylogram

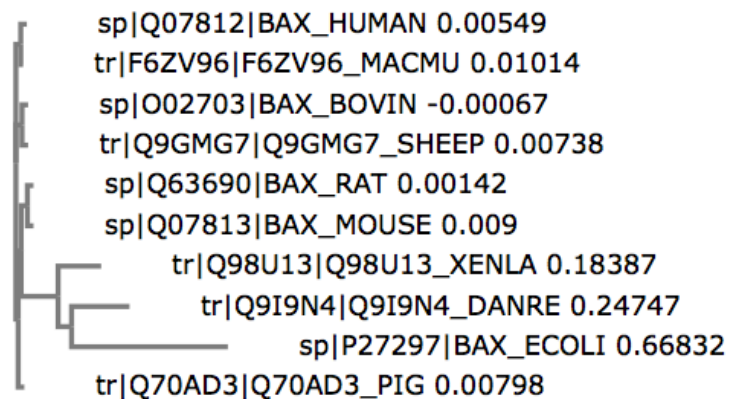
Branch length: ☒ Cladogram ☐ Real



Phylogram

Branch length: ☐ Cladogram ☒ Real

Real
(distance)



WE3

D. View conserved sequences in UCSC Genome Browser (The “conservome”)

The screenshot displays the UCSC Genome Browser interface. At the top, the navigation bar includes links for Home, Genomes, Blat, Tables, Gene Sorter, PCR, Session, FAQ, and Help. The main heading is "Human (*Homo sapiens*) Genome Browser Gateway". Below this, a search bar allows users to specify the clade (Mammal), genome (Human), assembly (Feb. 2009 (GRCh37/hg19)), and position or search term (MITF). A "submit" button is present. A link "Click here to reset the browser user interface settings to their defaults." is also visible.

Below the search bar, the "UCSC Genes" section lists several MITF transcripts, including MITF (uc021xam.1) at chr3:69985751-70017488. A callout box points to the "MITF" search term and says "Click to reset tracks and search for 'MITF'".

The main view shows the "UCSC Genome Browser on Human Feb. 2009 (GRCh37/hg19) Assembly". The search results show the MITF gene on chromosome 3, with coordinates chr3:69,985,751-70,017,488 (31,738 bp). The browser displays various tracks, including RefSeq Genes, Human mRNAs, Spliced ESTs, Layered H3K27ac, DNase Clusters, Transcription Factor ChIP-seq, and Human Cons. A callout box points to the "Human Cons" track and says "Scroll down to Comparative Genomics track and click 'Conservation'".

The "Comparative Genomics" section is shown at the bottom, with a "Conservation" track selected. The track shows conservation scores for various species, including Human, Mouse, Dog, Elephant, Opossum, Chicken, X_tropicalis, and Zebrafish. The track is currently set to "full" and "hide" other tracks like Indels, GERP, Evo Cpg, Primate Chain/Net, and Placental Chain/Net.

Home Genomes Genome Browser Blat Tables Gene Sorter PCR Session FAQ Help

Conservation Track Settings Subtracks! Description!

Vertebrate Multiz Alignment & Conservation (46 Species) [\(All Comparative Genomics tracks\)](#)

Maximum display mode: full [Reset to defaults](#)

Select views (help):

[Multiz Alignments](#) [Basewise Conservation \(phyloP\)](#) [Element Conservation \(phastCons\)](#) [Conserved Elements](#)

Multiz Alignments Configuration

Species selection: Defaults

Primate

☐ chimp ☐ gorilla ☐ orangutan ☒ rhesus ☐ baboon
☐ marmoset ☐ tarsier ☐ mouse lemur ☐ bushbaby

Placental Mammal

☐ tree shrew ☒ mouse ☐ rat ☐ kangaroo rat ☐ guinea pig
☐ squirrel ☐ rabbit ☐ pika ☐ alpaca ☐ dolphin
☐ cow ☐ horse ☐ cat ☒ dog ☐ microbat
☐ megabat ☐ hedgehog ☐ shrew ☒ elephant ☐ rock hyrax
☐ tenrec ☐ armadillo ☐ sloth

Vertebrate

☐ wallaby ☒ opossum
☐ lizard ☒ x. tropicalis
☐ medaka ☒ zebrafish

Multiple alignment base-
☐ Display bases identical
☒ Display chains between

Codon Translation:
 Default species to establish
☐ No codon translation
☒ Use default species read
☐ Use reading frames for
☐ Use reading frames for

UCSC Genome Browser on Human Feb. 2009 (GRCh37/hg19) Assembly

move <<< << < > >> >>> zoom in 1.5x 3x 10x base zoom out 1.5x 3x 10x

chr3:69,985,751-70,017,488 31,738 bp. enter position, gene symbol or search terms go

chr3 (p13) chr3:69,985,751-70,017,488 chr3 (p13)

Scale chr3: 69,985,000 69,995,000 70,005,000 70,015,000 hg19

UCSC Genes (RefSeq, UniProt, CCDS, Rfam, tRNAs & Comparative Genomics)

RefSeq Genes

Human mRNAs

Human mRNAs from GenBank

Spliced ESTs

Human ESTs That Have Been Spliced

HSK279C Mark: (Often Found Near Active Regulatory Elements) on 7 cell lines from ENCODE

Digital DNaseI Hypersensitivity Clusters from ENCODE

Transcription Factor ChIP-seq from ENCODE

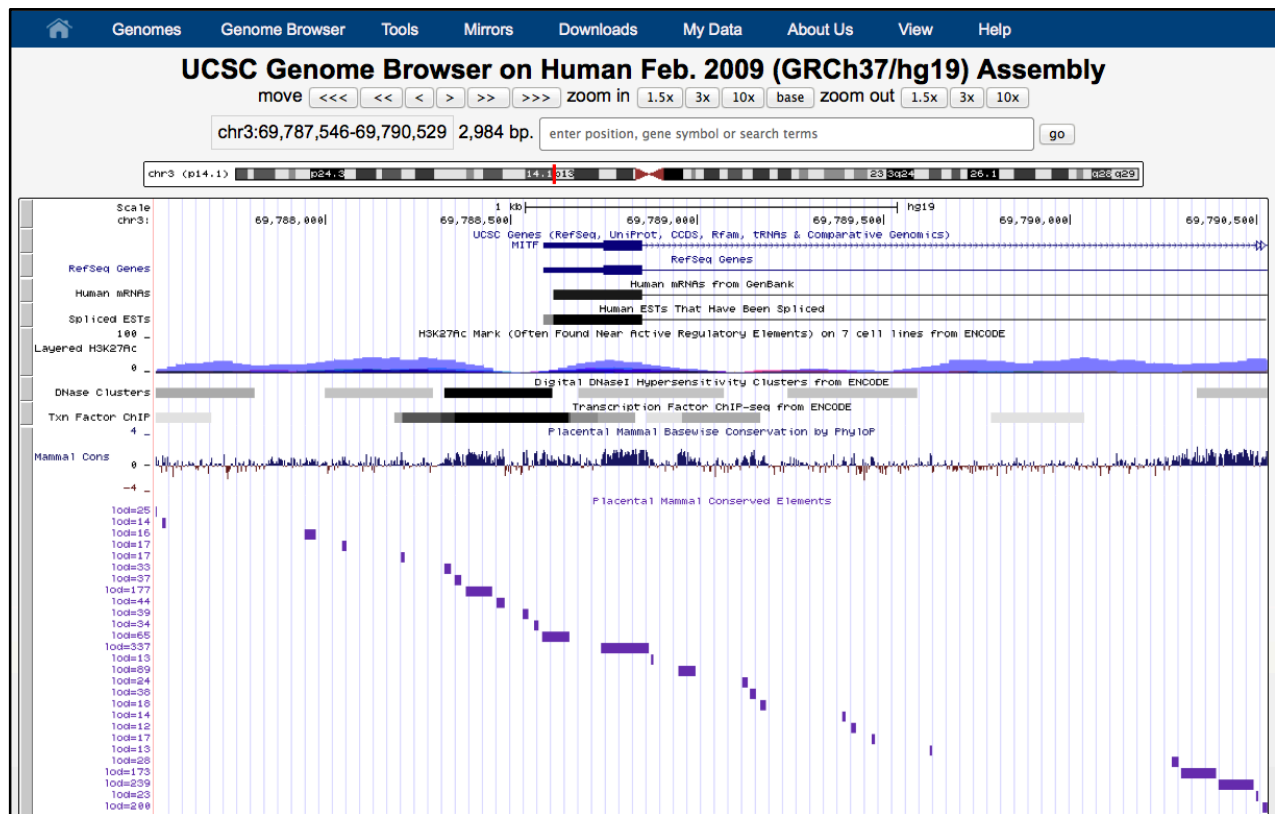
Placental Mammal Basewise Conservation by PhyloP

Placental Mammal Conserved Elements

Conserved regions of the human genome (based on analysis of multiple species)

Select "full" for conserved elements and push "submit".

WE3



Navigate to view the conserved elements at the 5' end the MITF gene.

Click to view sequence.

Placental Mammal Conserved Elements (lod=177)

Item: lod=177

Score: 539

Position: [chr3:69788379-69788447](#)

Band: 3p14.1

Genomic Size: 69

[View DNA for this feature](#) (hg19/Human)

[View table schema](#)

[Go to Conservation track controls](#)

Data last updated: 2009-11-20

```
>hg19_phastConsElements46wayPlacental_lod=177 range=chr3:69788379-69788447 5'pad=0 3'pad=0 strand=+ repeatMasking=none
GGGAGAGGAGCTGCTGACATTGACAATGAATCCAAACAGGAGTTGCACTA
GCGGTGTCCACCACGTTGC
```

WE3

E. Layer on transcription factor binding sites using ECR Browser

NCBI DCODE.org Comparative Genomics Developments
comparing genomes to decipher the code of gene regulation

Tools

- ECR Browser
- ECRbase
- Mulan
- zPicture
- eShadow
- CLARE
- DIRE
- SynoR
- Array2BIO
- multiTF
- rVista 2.0

Whole genome alignments

ECR Browser -- Evolutionary conservation of multiple genomes. Identification and sequence analysis of regulatory elements.
Genome Alignment in ECR Browser -- Align your FASTA nucleotide sequence to a genome of choice.

Multiple and pairwise sequence alignments

Mulan -- Full multiple sequence alignment. [Interactive conservation profiles, phylogenetic tree]
zPicture -- Stacked pairwise and multiple sequence alignment.
eShadow -- Phylogenetic shadowing of closely related species.

Regulation of co-expressed genes

CLARE -- Prediction of regulatory motifs and motif combinations in a set of co-functional enhancers.
DIRE -- Identification of proximal and Distant Regulatory Elements of co-regulated genes.
SynoR -- Prediction of synonymous regulatory elements in vertebrate genomes.

Identification of conserved transcription factor binding sites (cTFBS)

Search for "MITF" in ECR browser

base genome: Human feature or position (chrN:from-to): chr3:69,975,241-69,976,014 Submit

Genome Alignment in ECR Browser :: Align your sequence to a genome

[Instructions](#) on how to use ECR Browser.

Add ECR Browser search button to your Google Toolbar.
Click [HERE](#) and select your favorite genome.

Google Search

(You might need Administrative privileges to complete the installation.)

Features

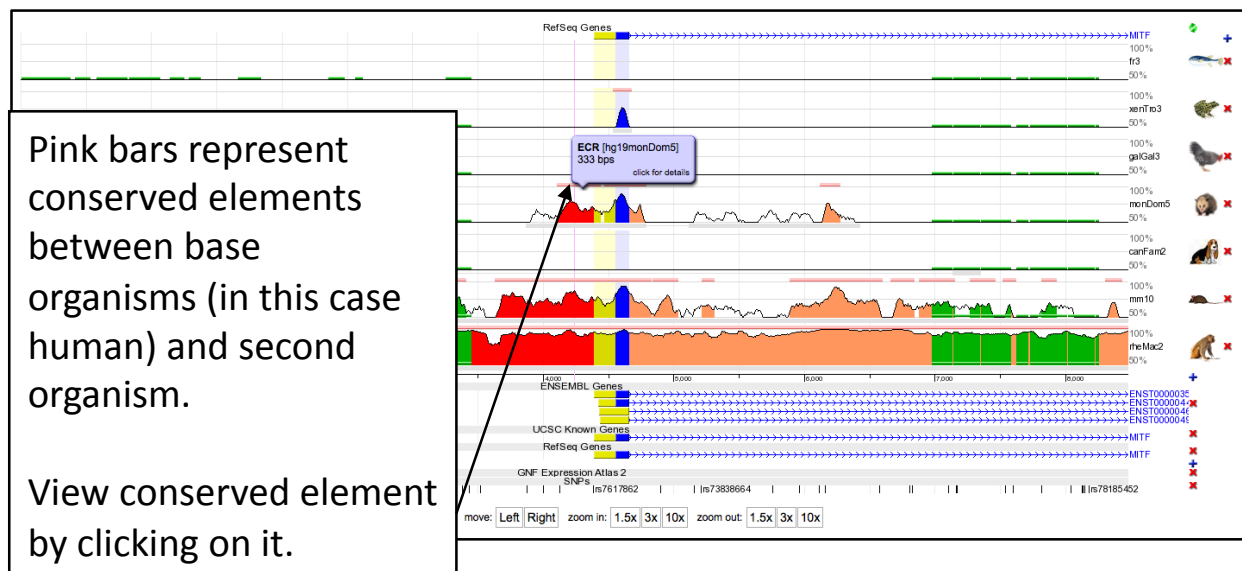
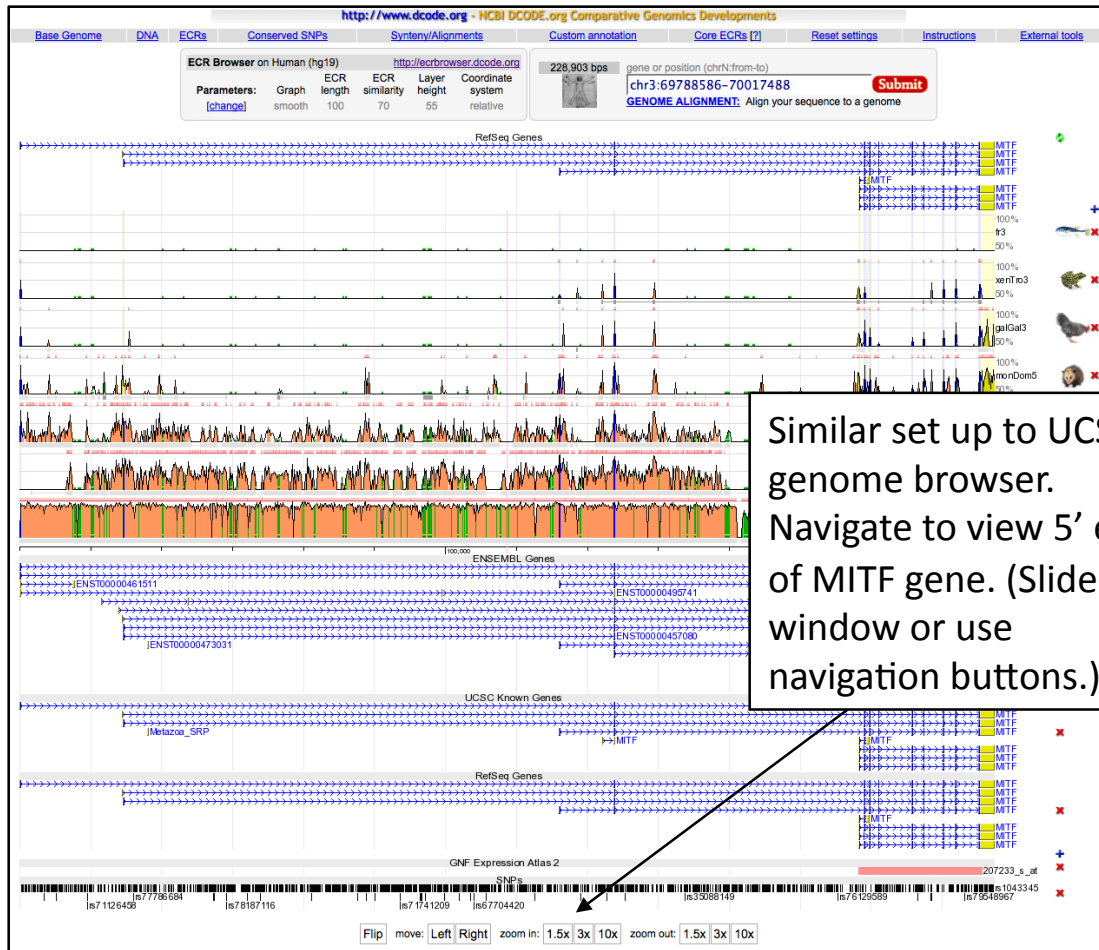
- multiple gene annotation tracks
- possibility to submit your own custom genome annotation
- share your custom annotation by submitting it to the [ECR Browser User Annotation Database](#)
- enhanced search through RefSeq, UCSC, Ensembl, mRNA, STS, SNPs, etc.
- zoom in/out using mouse wheel (similar to Google Maps)
- drag & drop rearrangement of gene annotation tracks
- drag & drop recentering of the conservation plot
- gene annotation drop on the conservation plot changes the reference annotation
- new alignments with repetitive elements included (colored in green)
- synteny annotation under each conservation track
- on-the-fly ECR and gene annotation
- single mouse-click *Grab ECR* function
- keyboard shortcuts:
 - o - zoom out 3x, l - zoom in 3x, > - shift to the right, < - shift to the left, f - flip the plot, g - genome selection window, p - parameters window, c - highlight coreECRs, r - reset parameters to defaults, f - refresh the page, a - additional alignments, m - main gene annotation, z - blastz-based genome alignment

Note: ECR Browser was tested on Internet Explorer 7 and Mozilla Firefox 2. We were also told it works on both Safari and Opera. Please update your Internet browser, if you experience unexpected behavior of ECR Browser.

RefSeq genes...

chr3:69985751-69988216	MITF	microphthalmia-associated transcription factor isoform 8
chr3:69812707-70017488	MITF	microphthalmia-associated transcription factor isoform 7
chr3:69812962-70017488	MITF	microphthalmia-associated transcription factor isoform 3
chr3:69985751-70017488	MITF	microphthalmia-associated transcription factor isoform 4
chr3:69985751-70017488	MITF	microphthalmia-associated transcription factor isoform 6
chr3:69915375-70017488	MITF	microphthalmia-associated transcription factor isoform 2
chr3:69788586-70017488	MITF	microphthalmia-associated transcription factor isoform 1
chr3:69985751-70017488	MITF	microphthalmia-associated transcription factor isoform 5

WE3



ECR :: Evolutionary Conserved Region Close

ECR [Evolutionary Conserved Region]
 location: chr3:69870755-69871673
 length: 919 bps
 identity: 69.5%

Alignment between human and opossum (Mondom)

conserved transcription factor binding sites (TFBS)

monDom4:chr6:175866007-175867093

175866007 175866027 175866047

monDom4 CAATAATATCTTCATCCCGAAATGCTGAAGGTTGTTTTCTCCTTGCCCTTTTTC

hg18 CAATAACGTTTTCATTACAGACCGAAGCAAGCTTCTTCCTGCTTC

69870755

RVISTA **DOWNLOAD**

Blastz alignment... ok
 Fetching annotation files... ok

Defining transcription factor binding sites

☒ TRANSFAC professional V10.2 library

Biological species

☒ vertebrates ☐ insects

☐ plants ☐ fungi

☐ nematodes ☐ bacteria

Matrix similarity

☒ Optimized for function

☐ Predefined as 0.85

Matrix selection

☐ use only high-specificity matrices

☐ User-defined consensus sequences

SUBMIT

The "optimized for function" option utilizes different cut-off parameters for different TFBS. It corresponds to "not more than 3 TFBS per 10kb of a random sequence" optimization [PMID: 15590941].

RVISTA **TRANSCRIPTION FACTORS**

Total number of transcription factor families: 467

SELECT TRANSCRIPTION FACTORS

SELECT SEPARATE TRANSCRIPTION FACTORS

A

☐ ACAA_B ☐ AFP1_Q6 ☐ AHR ☐ AHRARNT ☐ AHRHIF_Q6 ☐ AHR_Q5 ☐ AIRE ☐ ALPHACP1

☐ ALX4 ☐ AMEF2_Q6 ☐ AML1 ☐ AML1_Q6 ☐ AML_Q6 ☐ AP1 ☐ AP1F3_Q2 ☐ AP1_C

☐ AP1_Q2 ☐ AP1_Q4 ☐ AP1_Q6 ☐ AP2ALPHA ☐ AP2GAMMA ☐ AP2REP ☐ AP2_Q3 ☐ AP2_Q6

☐ AP3_Q6 ☐ AP4 ☐ AP4_Q5 ☐ AP4_Q6 ☐ APOLYA_B ☐ AR ☐ AREB6 ☐ ARNT

☐ ARP1 ☐ AR_Q2 ☐ AR_Q6 ☐ ATATA_B ☐ ATF ☐ ATF1_Q6 ☐ ATF3_Q6 ☐ ATF4_Q2

☐ ATF6 ☐ ATF_B

B .. C

☐ BACH1 ☐ BACH2 ☐ BARBIE ☐ BEL1_B ☐ BLIMP1_Q6 ☐ BRACH ☐ BRCA ☐ BRN2

☐ CAAT ☐ CAAT_C ☐ CACBINDING ☐ CACCBINDI ☐ CACD ☐ CAP ☐ CART1 ☐ CBF

☐ CDC5 ☐ CDP ☐ CDPCR1 ☐ CDPCR3 ☐ CDPCR3HD ☐ CDX2_Q5 ☐ CDXA ☐ CDX_Q5

☐ CEBP ☐ CEBPA ☐ CEBPB ☐ CEBPDELTA ☐ CEBPGAMMA ☐ CEBP_C ☐ CEBP_Q2 ☐ CEBP_Q3

☐ CETS168_Q6 ☐ CETS1P4 ☐ CHCH ☐ CHOP ☐ CHX10 ☐ CIZ ☐ CLOCKBMAL ☐ CLOX

☐ CMAF ☐ CMYB ☐ COMP1 ☐ COREBINDIN ☐ COUP ☐ COUPTF_Q6 ☐ COUP_DR1_Q ☐ CP2

☐ CREB ☐ CREBATF_Q5 ☐ CREBP1 ☐ CREBP1CIJUN ☐ CREBP1_Q2 ☐ CREB_Q2 ☐ CREB_Q3 ☐ CREB_Q4

☐ CREL ☐ CRX_Q4

SELECT ALL **RESET** **SUBMIT**

RVISTA

SUBMISSION RESULT

Data was successfully submitted for the rVista 2.0 processing and added to the queue
Reference index is [xbr04162012163407699](#)

CHECK IT

REGULATORY VISTA

RVISTA RESULTS

Request ID: [xbr11102014111323224](#)

Summary:
[5 conserved](#) and [5 aligned](#) transcription factor binding sites (TFBS) were identified

Dynamic visualization:
[Dynamically overlay](#) TFBS prediction with the conservation profile and perform clustering

Alignment:
[Highlight](#) TFBS positions in the alignment

```

      40          50
1ATAAGAGATAATAATCTATT
::| |||||::|||::
3CT--GAGATAATACTAAG
      60
  
```

Binding sites in the input sequences:
[138 TFBS](#) detected in the base sequence
[95 TFBS](#) detected in the second sequence

Input files:
Seq
Gen

RVISTA VISUALIZATION & CLUSTERING

Picture
Bases per layer: 1kb
Picture width (in pixels): 800
☐ Smooth plot

Clustering
☒ Individual clustering
☐ Combinatorial clustering
1 site(s) per 100 bps

Show
☒ conserved
☐ aligned
☐ all

[Select TF Subset](#) [Summary page](#) [List clustered TFBS](#)

MITF CONSERVED

The visualization shows a red bar representing the conservation profile across a genomic region. A green box highlights a specific binding site. The y-axis indicates the percentage of sequences containing the motif (0% to 100%).

REGULATORY VISTA

CONSERVED

0 sites

Site	Sequence	Score
1	V\$RFX_Q6 + 140-148 acGTTGCct 161-169 acGTTGCct	95.00
2	V\$RFXI_01 - 140-156 acGTTGCCTCTCCCCcg 161-177 acGTTGCCTCGAGGCcg	80.00
3	V\$STAF_02 - 154-174 ccGCCGTGGCCTTCTGGGAgt 175-195 ccGTCGCTCTGCTGGGAAtt	80.00
4	V\$MTF1_Q4 + 245-258 gcGGCGCTCGGCcg 267-280 ggAGCGCCCGGCcg	85.00
5	V\$ZF5_01 + 248-255 GCGCTCg 270-277 GCGCCGcg	85.00

