

Bioinformatika és genomanalízis az orvostudományban

Integrált biológiai adatbázisok

Cserző Miklós

2018

A mai előadás

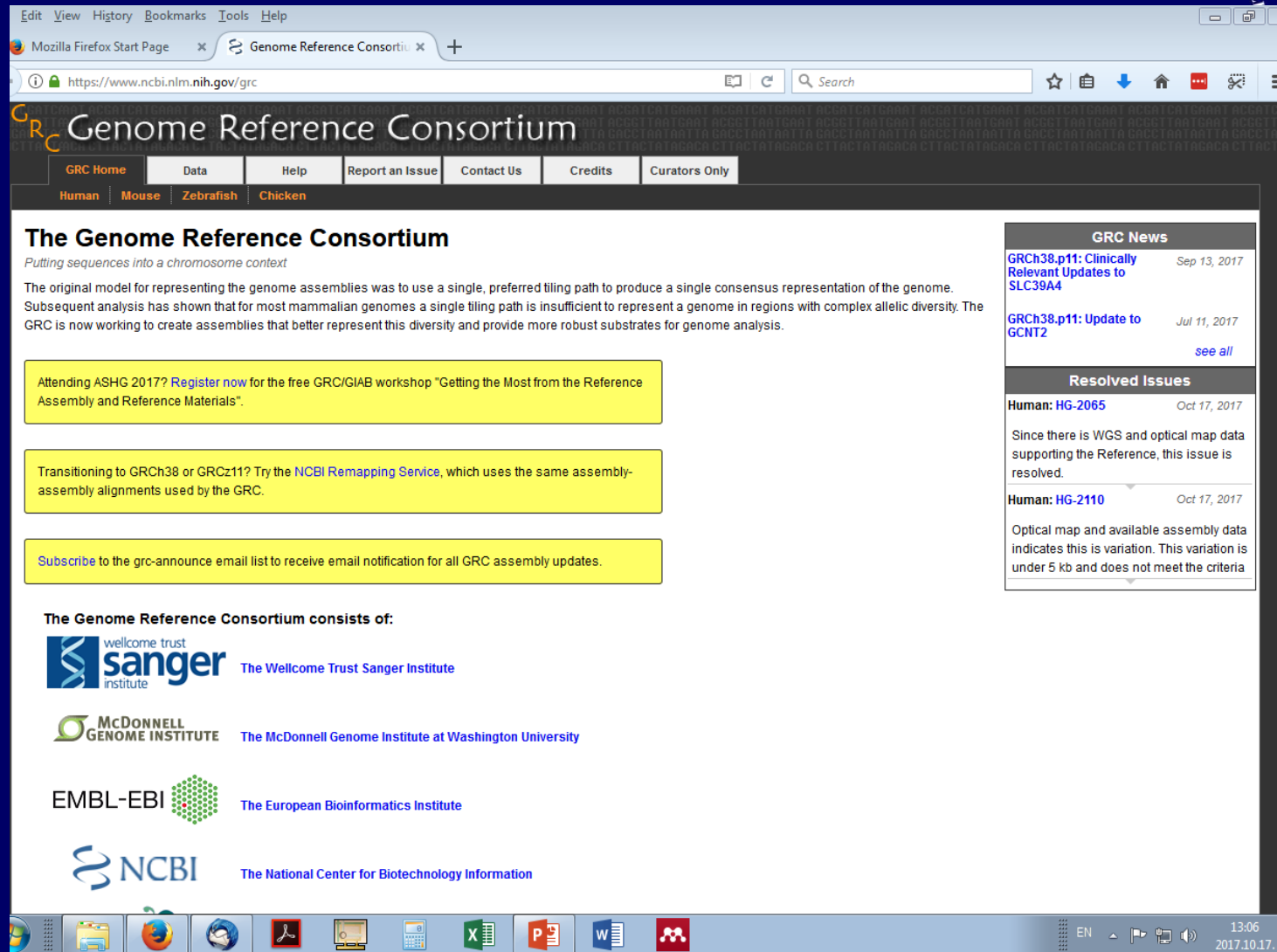
- A genom annotálás jelentősége
- Genome Reference Consortium
- Gene Ontology
- Az „ensembl” pipeline
- Lekérdezés a web-felületen keresztül
- Az osztott annotációs rendszer (DAS)
- A BioMart felület használata
- Adatelérés FTP-portálon

Miért annotálunk genomot?

- A szekvenálás gyors és olcsó
- Az adatok kísérletes igazolása viszont drága
- Az élőlények származástani rokonságban állnak egymással
- Az egyik élő rendszerben megszerzett kísérletes adatot fel lehet használni egy rokon esetében is
- Ezzel időt, pénzt, fáradságot lehet megtakarítani

Genom Reference Consortium

- A nagy genom szekvenálások nemzetközi összefogással mennek
- Nemcsak egyetlen egyed genomja érdekes
- Kell egy referencia az egyes fajokra
- A fajon belüli eltéréseket ehhez viszonyítjuk
- Honlap:
<http://www.ncbi.nlm.nih.gov/projects/genome/assembly/grc/index.shtml>



The Genome Reference Consortium
Putting sequences into a chromosome context

The original model for representing the genome assemblies was to use a single, preferred tiling path to produce a single consensus representation of the genome. Subsequent analysis has shown that for most mammalian genomes a single tiling path is insufficient to represent a genome in regions with complex allelic diversity. The GRC is now working to create assemblies that better represent this diversity and provide more robust substrates for genome analysis.

Attending ASHG 2017? [Register now](#) for the free GRC/GIAB workshop "Getting the Most from the Reference Assembly and Reference Materials".

Transitioning to GRCh38 or GRCh11? Try the [NCBI Remapping Service](#), which uses the same assembly-assembly alignments used by the GRC.

[Subscribe](#) to the grc-announce email list to receive email notification for all GRC assembly updates.

The Genome Reference Consortium consists of:

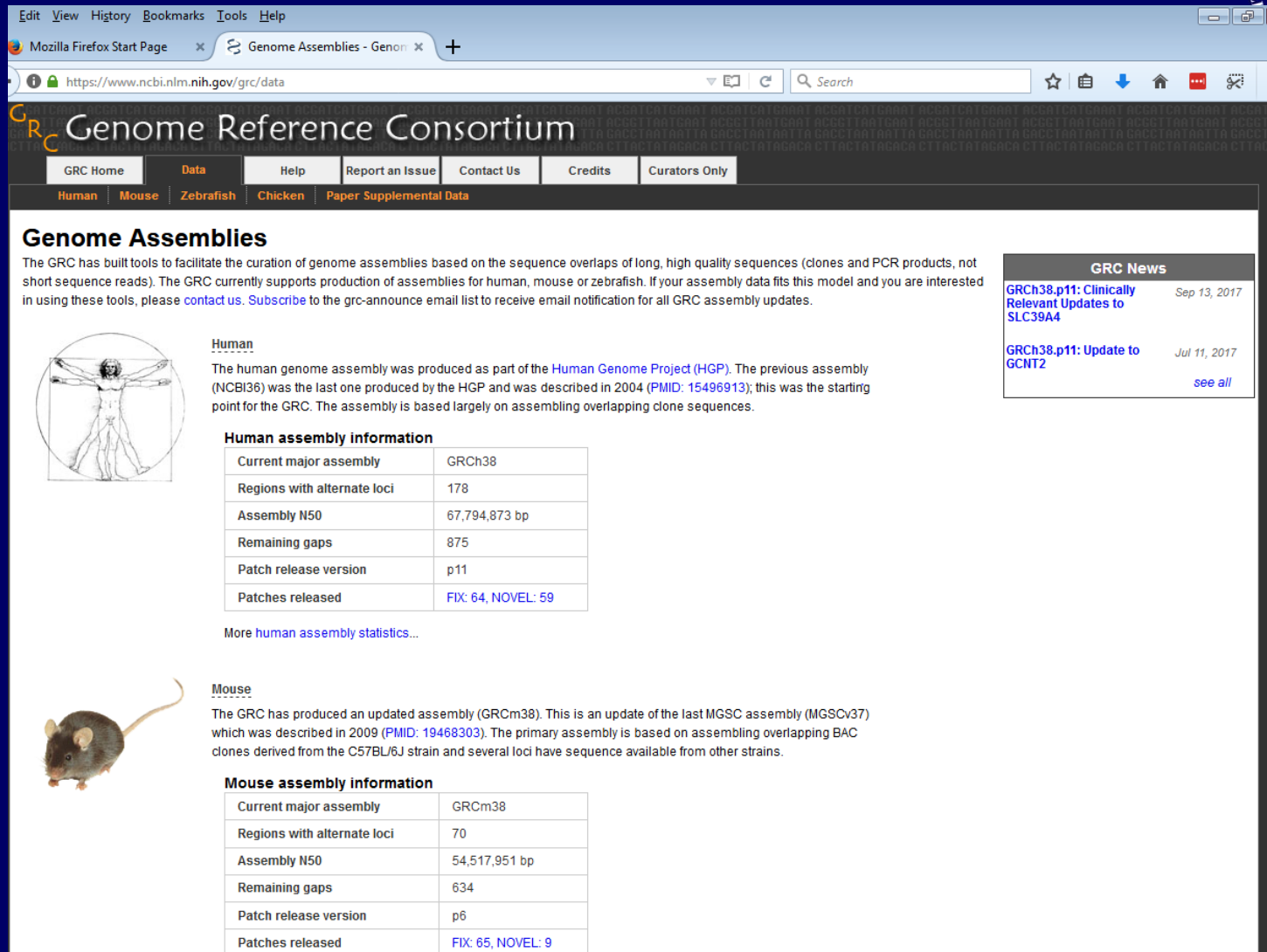
-  **The Wellcome Trust Sanger Institute**
-  **The McDonnell Genome Institute at Washington University**
-  **The European Bioinformatics Institute**
-  **The National Center for Biotechnology Information**

GRC News

- GRCh38.p11: Clinically Relevant Updates to SLC39A4** Sep 13, 2017
- GRCh38.p11: Update to GCNT2** Jul 11, 2017 [see all](#)

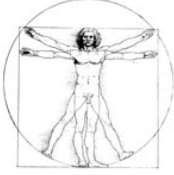
Resolved Issues

- Human: HG-2065** Oct 17, 2017
Since there is WGS and optical map data supporting the Reference, this issue is resolved.
- Human: HG-2110** Oct 17, 2017
Optical map and available assembly data indicates this is variation. This variation is under 5 kb and does not meet the criteria



Genome Assemblies

The GRC has built tools to facilitate the curation of genome assemblies based on the sequence overlaps of long, high quality sequences (clones and PCR products, not short sequence reads). The GRC currently supports production of assemblies for human, mouse or zebrafish. If your assembly data fits this model and you are interested in using these tools, please [contact us](#). [Subscribe](#) to the grc-announce email list to receive email notification for all GRC assembly updates.



Human

The human genome assembly was produced as part of the [Human Genome Project \(HGP\)](#). The previous assembly (NCBI36) was the last one produced by the HGP and was described in 2004 (PMID: [15496913](#)); this was the starting point for the GRC. The assembly is based largely on assembling overlapping clone sequences.

Human assembly information

Current major assembly	GRCh38
Regions with alternate loci	178
Assembly N50	67,794,873 bp
Remaining gaps	875
Patch release version	p11
Patches released	FIX: 64, NOVEL: 59

[More human assembly statistics...](#)



Mouse

The GRC has produced an updated assembly (GRCm38). This is an update of the last MGSC assembly (MGSCv37) which was described in 2009 (PMID: [19468303](#)). The primary assembly is based on assembling overlapping BAC clones derived from the C57BL/6J strain and several loci have sequence available from other strains.

Mouse assembly information

Current major assembly	GRCm38
Regions with alternate loci	70
Assembly N50	54,517,951 bp
Remaining gaps	634
Patch release version	p6
Patches released	FIX: 65, NOVEL: 9

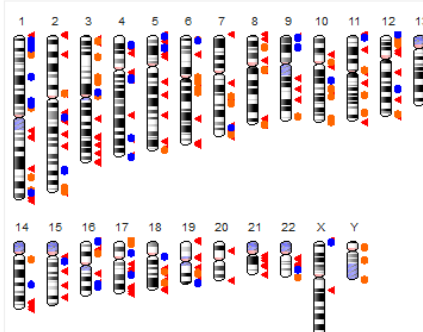
GRC News

[GRCh38.p11: Clinically Relevant Updates to SLC39A4](#) Sep 13, 2017

[GRCh38.p11: Update to GCNT2](#) Jul 11, 2017 [see all](#)

Human Genome Overview

Information about the continuing improvement of the human genome



Legend:

- Region containing alternate loci
- Region containing fix patches
- Region containing novel patches

Ideogram of the latest human assembly, GRCh38.p11

The GRC is working hard to provide the best possible reference assembly for human. We do this by both generating multiple representations (**alternate loci**) for regions that are too complex to be represented by a single path. Additionally, we are releasing regional fixes known as **patches**. This allows users who are interested in a specific locus to get an improved representation without affecting users who need chromosome coordinate stability.

Download data:

- GRCh38.p11 (latest minor release) FTP
- GRCh38 (latest major release) FTP
- Genomic regions under review FTP
- Current Tiling Path Files (TPFs)

Transitioning to GRCh38? Try the NCBI Remapping Service, which uses the same assembly-assembly alignments used by the GRC.

Next assembly update
The next assembly update (GRCh38.p12) will be a minor (patch) release in winter 2017.

GRC News

- GRCh38.p11: Clinically Relevant Updates to SLC39A4** Sep 13, 2017
- GRCh38.p11: Update to GCNT2** Jul 11, 2017 [see all](#)

Resolved Human Issues

- HG-2065** Oct 17, 2017
Since there is WGS and optical map data supporting the Reference, this issue is resolved.
- HG-2110** Oct 17, 2017
Optical map and available assembly data indicates this is variation. This variation is under 5 kb and does not meet the criteria [see all](#)

GRCh38.p11 GRCh37.p13 GRCh37

GRCh38.p11

Release date: June 14, 2017
Release type: minor
Release notes: GRCh38.p11 is the eleventh patch release for the GRCh38 reference assembly. No chromosome coordinates changed. This release includes 11 FIX patches and 10 NOVEL patch. The total number of patch scaffolds is now: 64 FIX and 59 NOVEL.
Assembly accessions: GenBank: [GCA_000001405.26](#), RefSeq: [GCF_000001405.37](#)

Pseudoautosomal regions

Human Genome Overview - Human Genome Overview - +

https://www.ncbi.nlm.nih.gov/grc/human

Ideogram of the latest human assembly, GRCh38.p11

GRCh38.p11 GRCh37.p13 GRCh37

GRCh38.p11

Release date: June 14, 2017
Release type: minor
Release notes: GRCh38.p11 is the eleventh patch release for the GRCh38 reference assembly. No chromosome coordinates changed. This release includes 11 FIX patches and 10 NOVEL patch. The total number of patch scaffolds is now: 64 FIX and 59 NOVEL.
Assembly accessions: GenBank: [GCA_000001405.26](#), RefSeq: [GCF_000001405.37](#)

Pseudoautosomal regions

Name	Chr	Start	Stop
PAR#1	X	10,001	2,781,479
PAR#2	X	155,701,383	156,030,895
PAR#1	Y	10,001	2,781,479
PAR#2	Y	56,887,903	57,217,415

Modeled centromeres and heterochromatin regions [Download](#)

Items 1 - 5 of 25 << First < Prev Page 1 of 5 Next > Last >>

Name	Chr	Start	Stop	Length
CEN1	1	122,026,460	125,184,587	3,158,128
CEN2	2	92,188,146	94,090,557	1,902,412
CEN3	3	90,772,459	93,655,574	2,883,116
CEN4	4	49,708,101	51,743,951	2,035,851
CEN5	5	46,485,901	50,059,807	3,573,907

Items 1 - 5 of 25 << First < Prev Page 1 of 5 Next > Last >>

Patches and alternate loci [Download](#)

Filter by: Gene

Items 1 - 20 of 384 << First < Prev Page 1 of 20 Next > Last >>

Scaffold IDs	Assembly unit	Region name	Chr	Start	Stop	Type
K1270762.1 NT_187515.1	ALT_REF_LOCL_1	REGION108	1	2,448,811	2,791,270	alt
K1270766.1 NT_187517.1	ALT_REF_LOCL_1	PRAME_REGION_1	1	12,818,488	13,312,803	alt

Mivel annotáljunk?

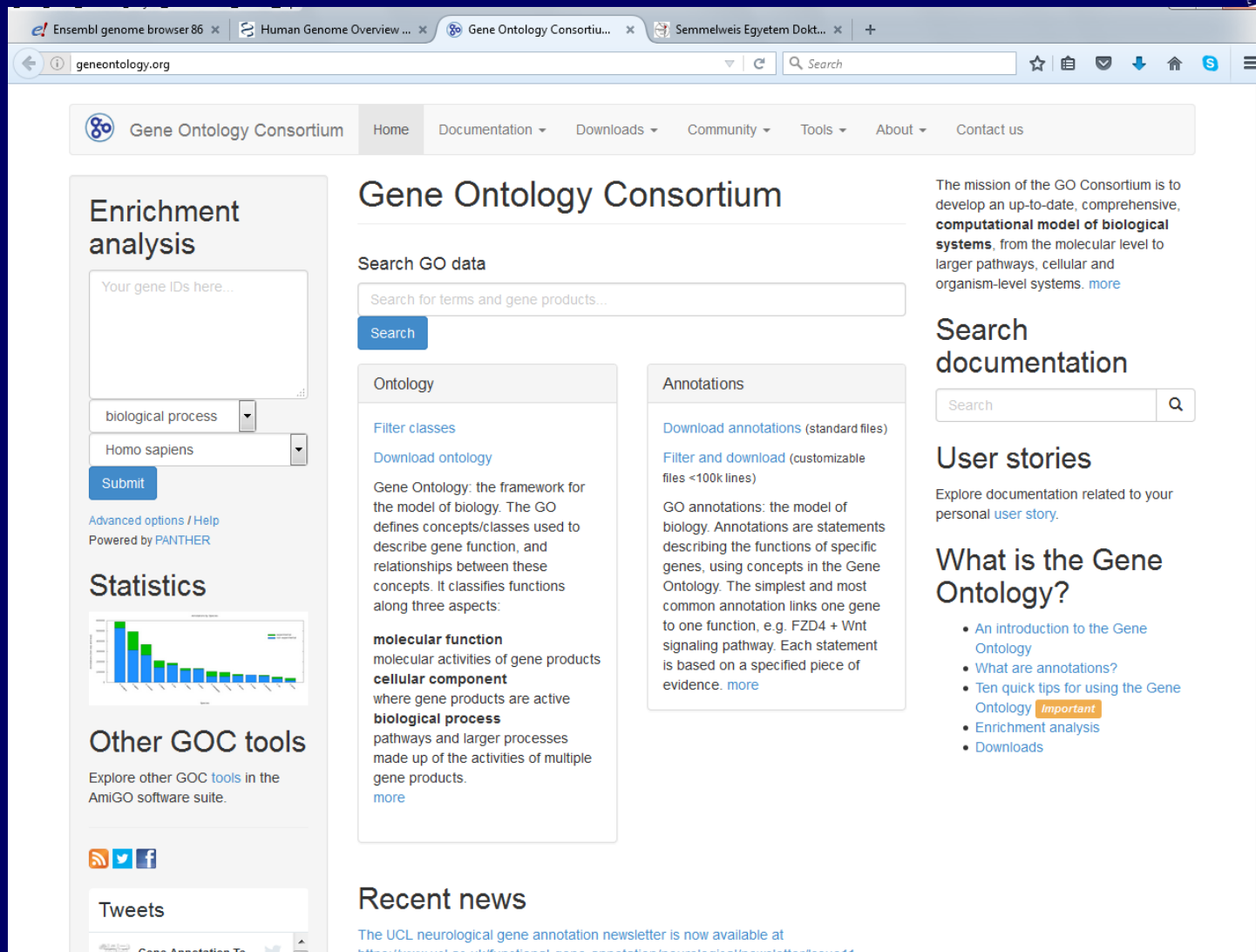
- Gene Ontology Consortium
- Zárt annotációs szótár – egy bizottság dönt, mi kerülhet bele
- Cél: a gének pontos és teljes leírása
- Követelmény: a leírás legyen koherens, általános és gépi feldolgozásra alkalmas

A megvalósítás

- Besorolás három szempont szerint:
 - Folyamat (biological_process)
 - Sejtalkotó (cellular_component)
 - Működés (molecular_function)
- Három viszony:
 - „is_a”
 - „part_of”
 - „regulates”

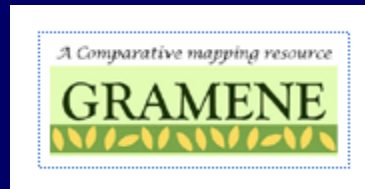
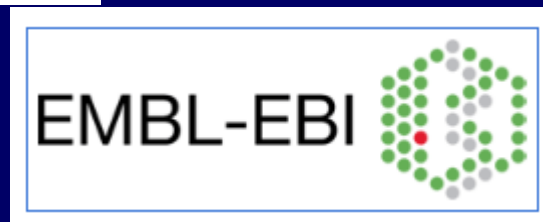
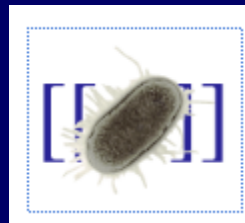
A módszer

- 12 modell-lény – és egy népes szakértői csapat
- Elsődleges forrás: közvetlen kísérletes adatok
- Másodlagos forrás: homológ fehérjék annotációja
- Azt is jelzik, ha nincs adat
- Honlap: <http://www.geneontology.org/>



The screenshot shows the Gene Ontology Consortium website. The browser tabs include 'Ensembl genome browser 86', 'Human Genome Overview ...', 'Gene Ontology Consortiu...', and 'Semmelweis Egyetem Dokt...'. The address bar shows 'geneontology.org'. The website features a navigation bar with links to Home, Documentation, Downloads, Community, Tools, About, and Contact us. The main content area is divided into several sections:

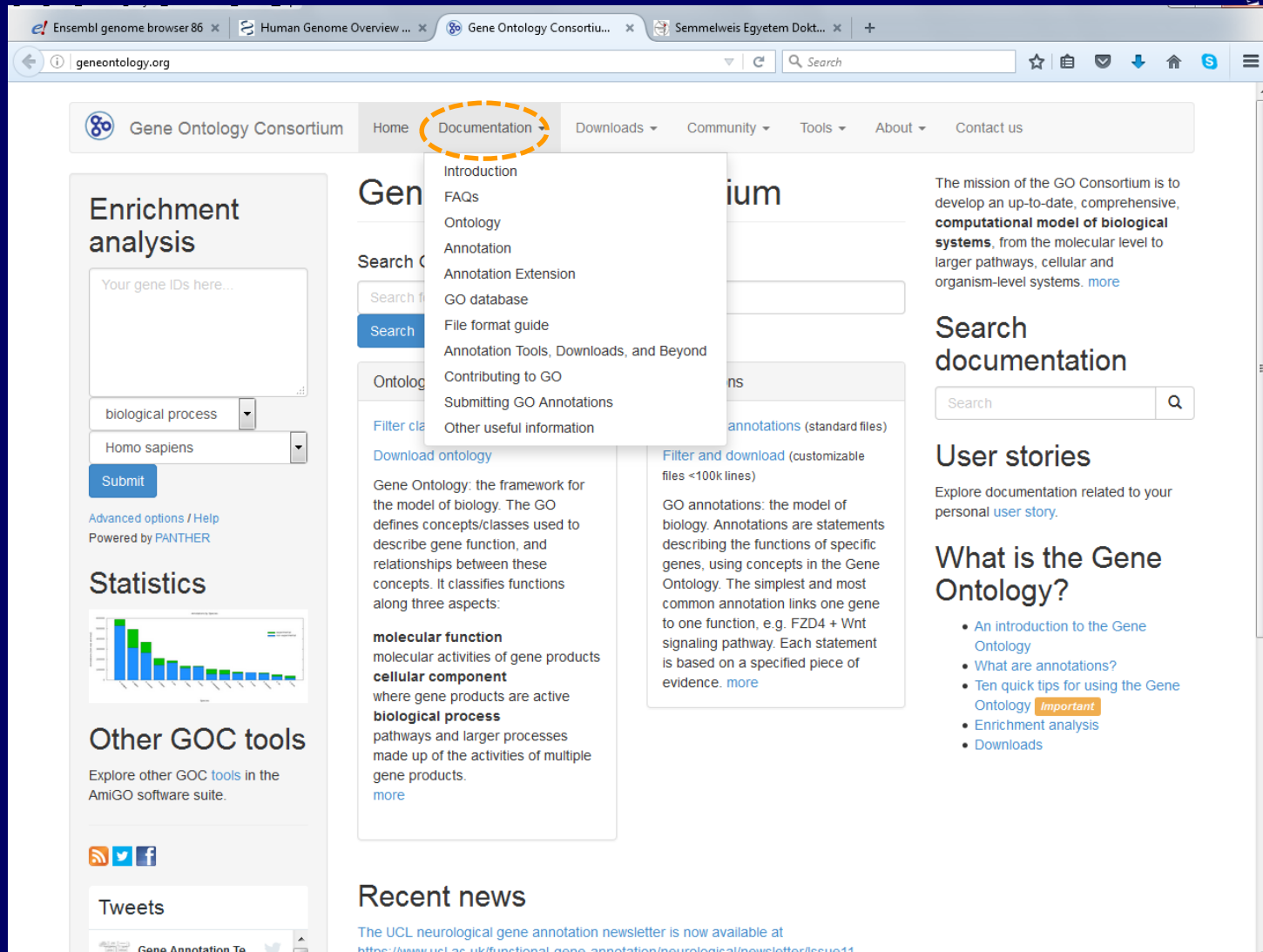
- Enrichment analysis:** A section for submitting gene IDs and selecting a biological process and organism (Homo sapiens).
- Statistics:** A bar chart showing the distribution of gene products across different biological processes.
- Other GOC tools:** A section for exploring other GOC tools in the AmiGO software suite.
- Gene Ontology Consortium:** A central section with a search bar and a description of the consortium's mission.
- Search GO data:** A section for searching for terms and gene products.
- Annotations:** A section for downloading annotations and filtering them.
- User stories:** A section for exploring documentation related to personal user stories.
- What is the Gene Ontology?:** A section with links to various resources, including an introduction to the Gene Ontology, what are annotations, ten quick tips for using the Gene Ontology, enrichment analysis, and downloads.
- Recent news:** A section for the latest news, including a link to the UCL neurological gene annotation newsletter.



Meeting Photo: GO Consortium Meeting

Hôtel Mont Gabriel, Sainte-Adèle, Québec, Canada, October 21-23, 2008

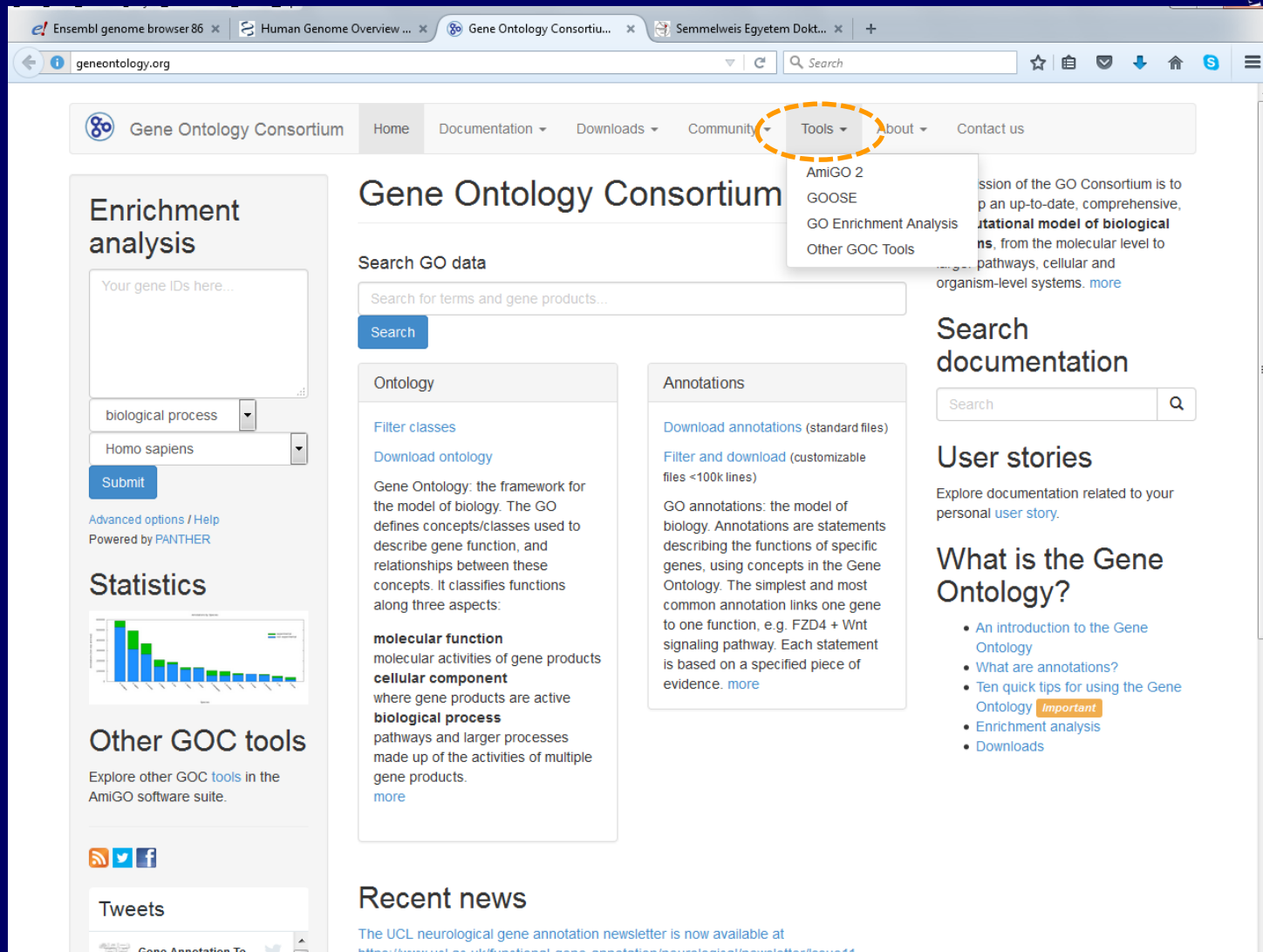




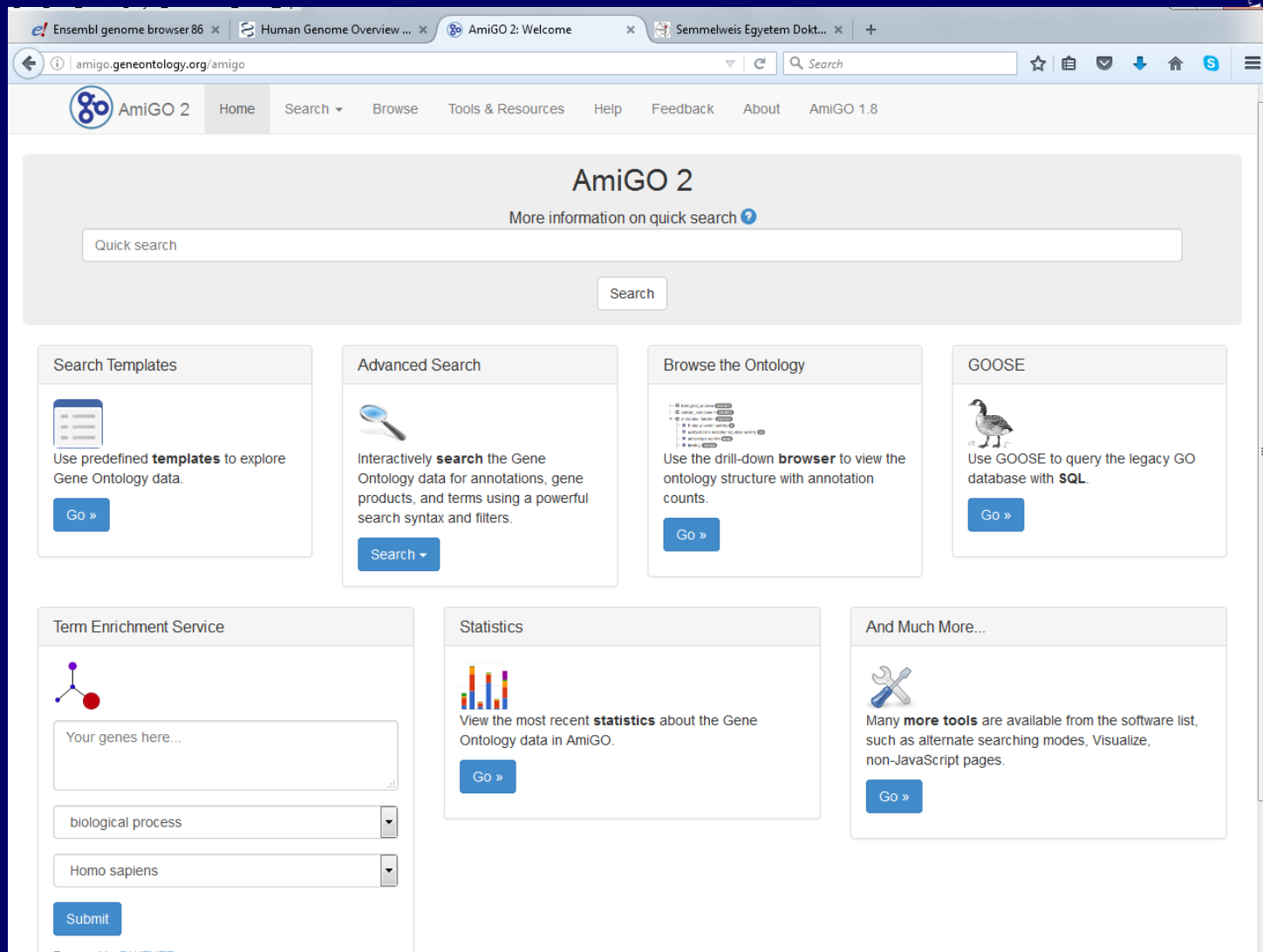
The screenshot shows the Gene Ontology Consortium website. The 'Documentation' menu is open, listing the following items:

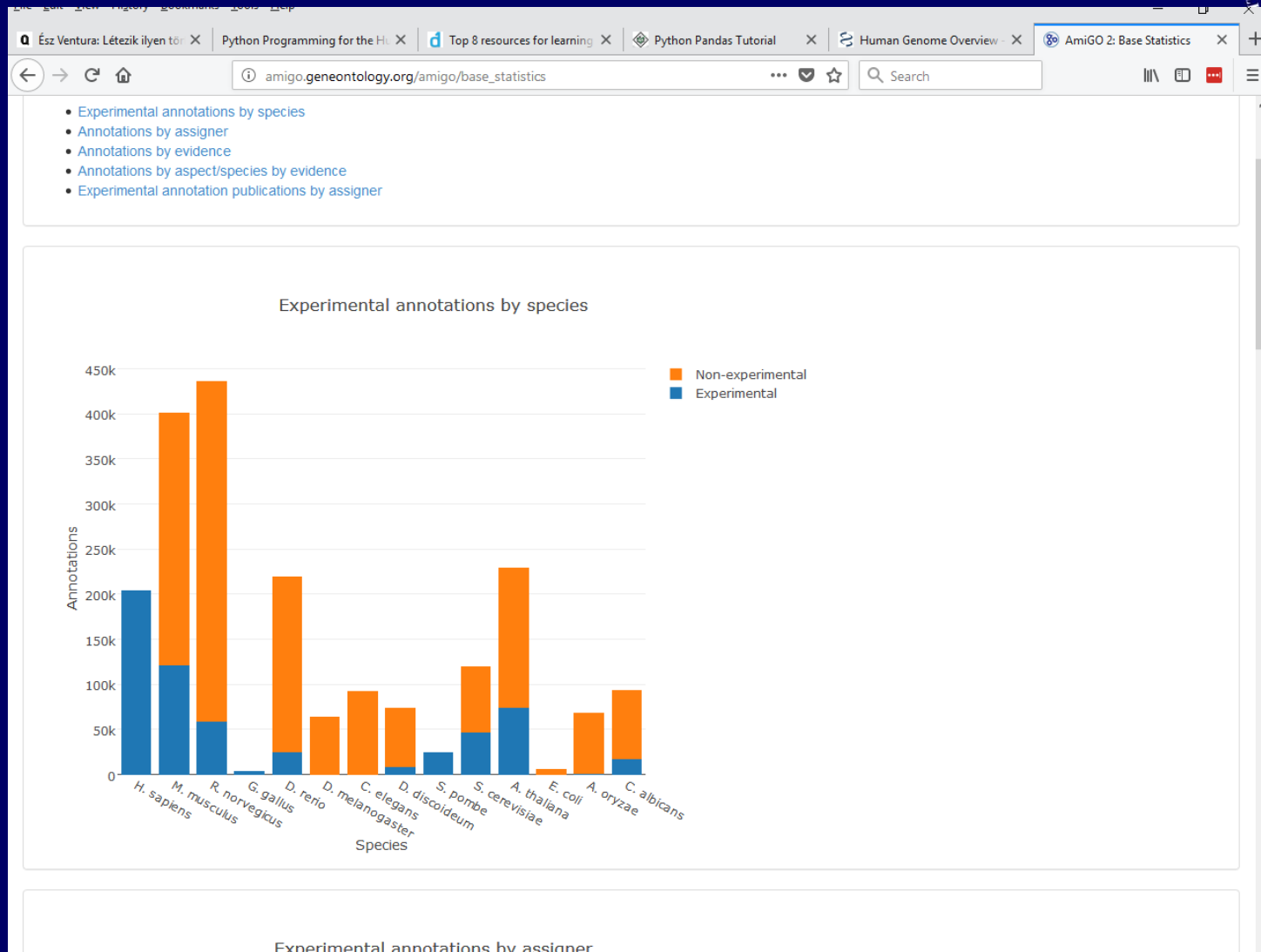
- Introduction
- FAQs
- Ontology
- Annotation
- Annotation Extension
- GO database
- File format guide
- Annotation Tools, Downloads, and Beyond
- Contributing to GO
- Submitting GO Annotations
- Other useful information

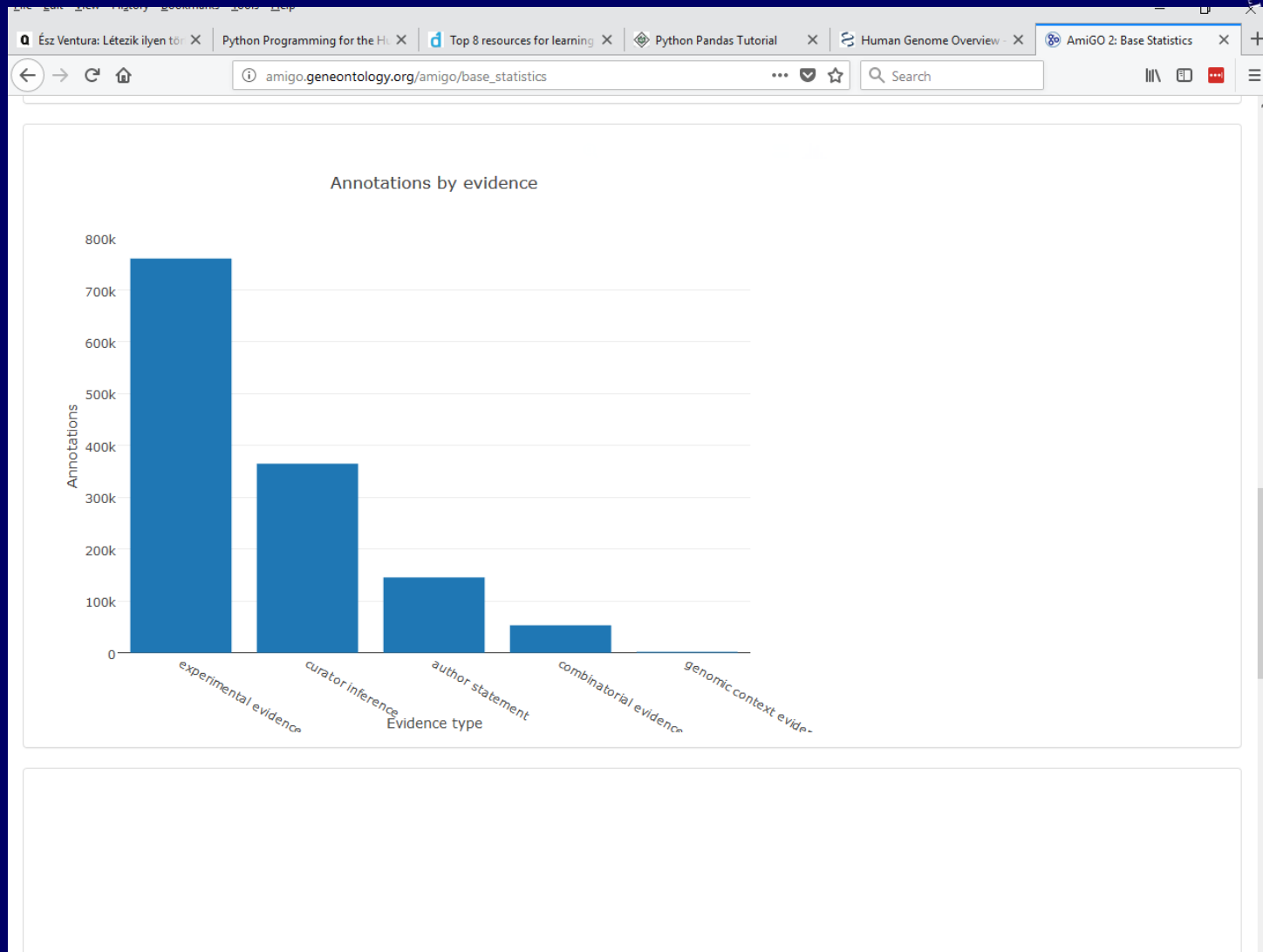
The website also features sections for Enrichment analysis, Statistics, Other GOC tools, Recent news, Search documentation, User stories, and What is the Gene Ontology?.

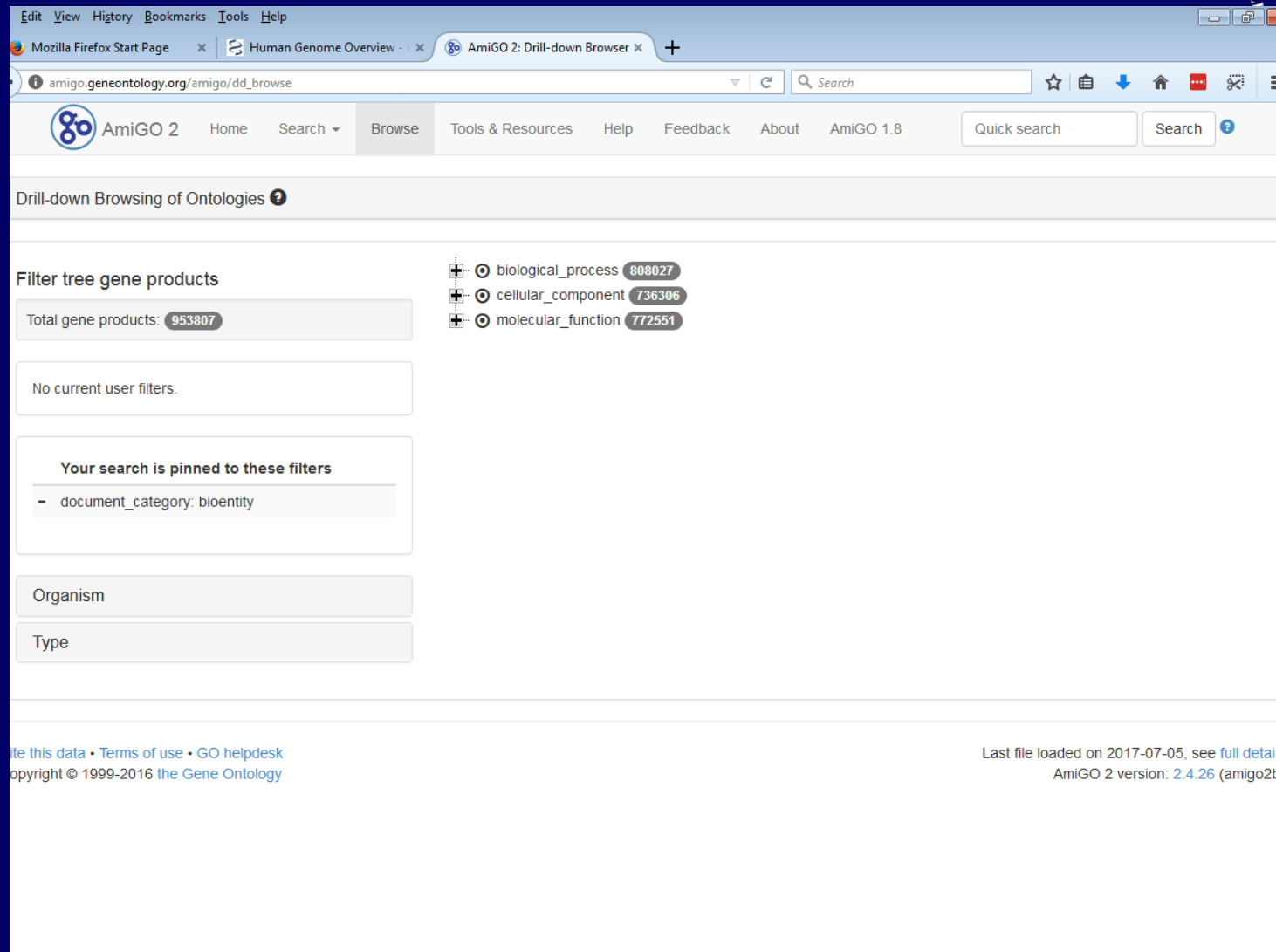


The screenshot shows the Gene Ontology Consortium website. The 'Tools' dropdown menu is open, showing options: AmiGO 2, GOOSE, GO Enrichment Analysis, and Other GOC Tools. The 'GO Enrichment Analysis' option is highlighted. The main content area includes sections for 'Enrichment analysis', 'Search GO data', 'Statistics', 'Other GOC tools', 'Recent news', 'Search documentation', 'User stories', and 'What is the Gene Ontology?'. The 'Enrichment analysis' section has a text input for 'Your gene IDs here...', a dropdown for 'biological process', and a dropdown for 'Homo sapiens'. The 'Search GO data' section has a search bar and a 'Search' button. The 'Statistics' section shows a bar chart. The 'Other GOC tools' section has a link to 'Explore other GOC tools in the AmiGO software suite'. The 'Recent news' section has a link to 'The UCL neurological gene annotation newsletter is now available at https://www.ucl.ac.uk/functional_gene_annotation/neurological/newsletter/issue11'. The 'Search documentation' section has a search bar. The 'User stories' section has a link to 'Explore documentation related to your personal user story'. The 'What is the Gene Ontology?' section has a list of links: 'An introduction to the Gene Ontology', 'What are annotations?', 'Ten quick tips for using the Gene Ontology', 'Enrichment analysis', and 'Downloads'.









The screenshot shows the AmiGO 2 Drill-down Browser interface in a Mozilla Firefox browser. The address bar shows the URL amigo.geneontology.org/amigo/dd_browse. The page title is "Drill-down Browsing of Ontologies". The main content area displays a filter tree for gene products. The "Filter tree gene products" section shows a total of 953807 gene products. The "Your search is pinned to these filters" section shows a filter for "document_category: bioentity". The "Organism" and "Type" sections are currently empty. The footer contains links for "Site this data", "Terms of use", and "GO helpdesk", along with copyright information for 1999-2016 the Gene Ontology. The last file loaded is dated 2017-07-05, and the AmiGO 2 version is 2.4.26.

Drill-down Browsing of Ontologies

Filter tree gene products

Total gene products: 953807

No current user filters.

Your search is pinned to these filters

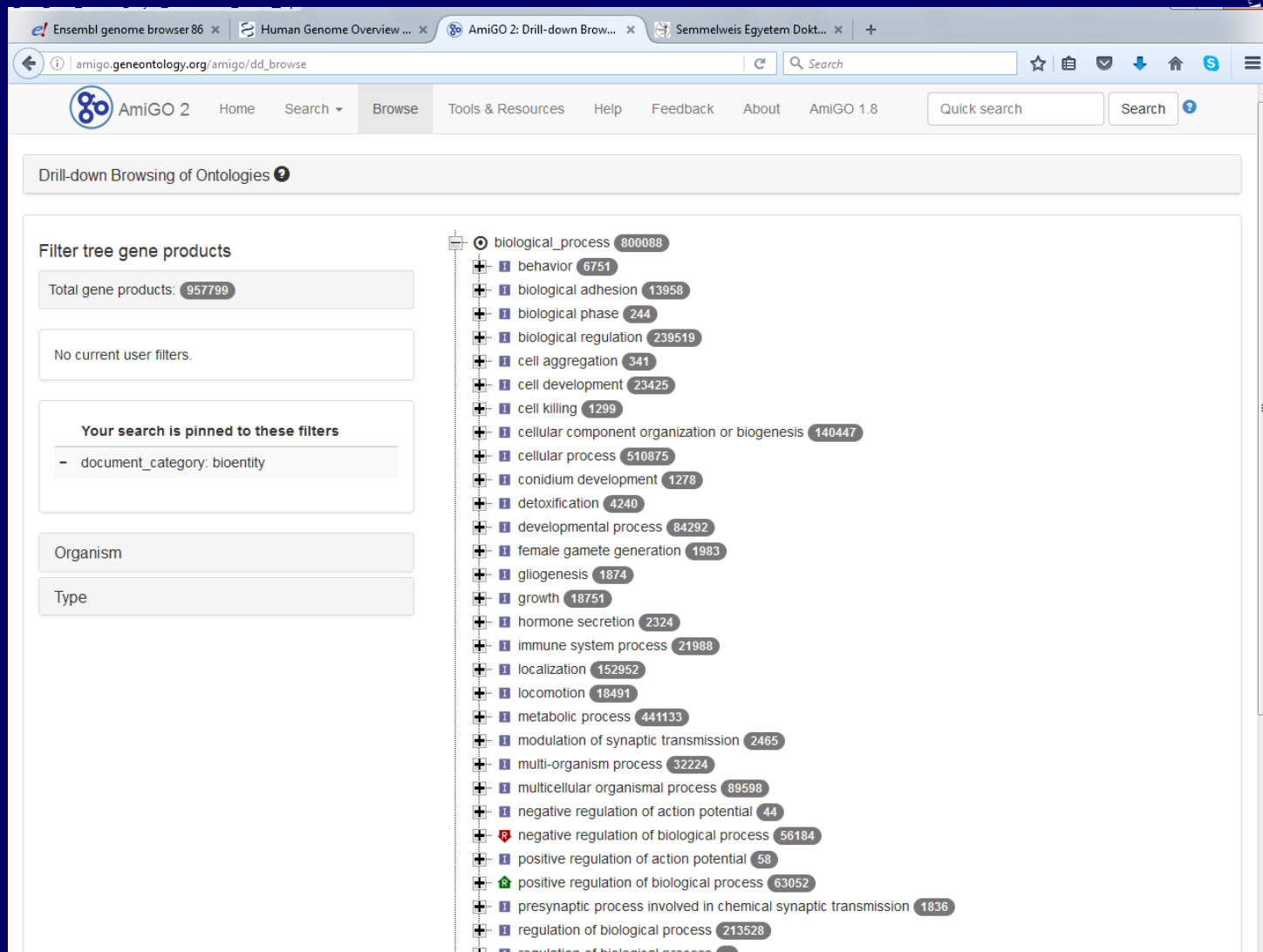
- document_category: bioentity

Organism

Type

Site this data • Terms of use • GO helpdesk
copyright © 1999-2016 the Gene Ontology

Last file loaded on 2017-07-05, see full details
AmiGO 2 version: 2.4.26 (amigo2)



Ensembl genome browser 86 x Human Genome Overview ... x AmiGO 2: Drill-down Brow... x Semmelweis Egyetem Dokt... x

amigo.geneontology.org/amigo/dd_browse

AmiGO 2 Home Search Browse Tools & Resources Help Feedback About AmiGO 1.8 Quick search Search

Drill-down Browsing of Ontologies

Filter tree gene products

Total gene products: 957799

No current user filters.

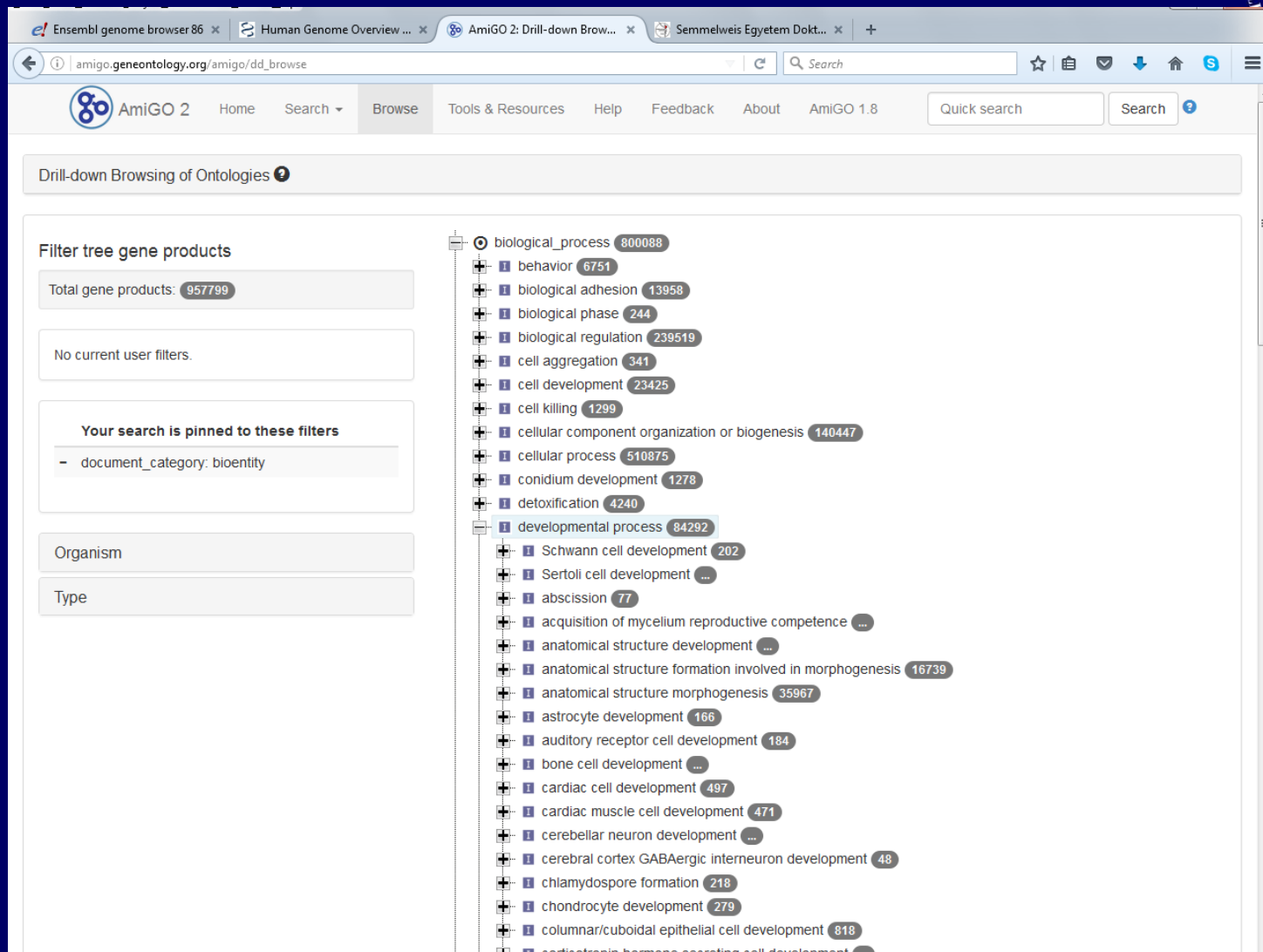
Your search is pinned to these filters

- document_category: bioentity

Organism

Type

- biological_process (800088)
 - behavior (6751)
 - biological adhesion (13958)
 - biological phase (244)
 - biological regulation (239519)
 - cell aggregation (341)
 - cell development (23425)
 - cell killing (1299)
 - cellular component organization or biogenesis (140447)
 - cellular process (510875)
 - conidium development (1278)
 - detoxification (4240)
 - developmental process (84292)
 - female gamete generation (1983)
 - gliogenesis (1874)
 - growth (18751)
 - hormone secretion (2324)
 - immune system process (21988)
 - localization (152952)
 - locomotion (18491)
 - metabolic process (441133)
 - modulation of synaptic transmission (2465)
 - multi-organism process (32224)
 - multicellular organismal process (89598)
 - negative regulation of action potential (44)
 - negative regulation of biological process (56184)
 - positive regulation of action potential (58)
 - positive regulation of biological process (63052)
 - presynaptic process involved in chemical synaptic transmission (1836)
 - regulation of biological process (213528)
 - regulation of biological process



Ensembl genome browser 86 x Human Genome Overview ... x AmiGO 2: Drill-down Brow... x Semmelweis Egyetem Dokt... x

amigo.geneontology.org/amigo/dd_browse

AmiGO 2 Home Search Browse Tools & Resources Help Feedback About AmiGO 1.8 Quick search Search

Drill-down Browsing of Ontologies

Filter tree gene products

Total gene products: 957799

No current user filters.

Your search is pinned to these filters

- document_category: bioentity

Organism

Type

- biological_process (800088)
 - behavior (6751)
 - biological adhesion (13958)
 - biological phase (244)
 - biological regulation (239519)
 - cell aggregation (341)
 - cell development (23425)
 - cell killing (1299)
 - cellular component organization or biogenesis (140447)
 - cellular process (510875)
 - conidium development (1278)
 - detoxification (4240)
 - developmental process (84292)
 - Schwann cell development (202)
 - Sertoli cell development (...)
 - abscission (77)
 - acquisition of mycelium reproductive competence (...)
 - anatomical structure development (...)
 - anatomical structure formation involved in morphogenesis (16739)
 - anatomical structure morphogenesis (35967)
 - astrocyte development (166)
 - auditory receptor cell development (184)
 - bone cell development (...)
 - cardiac cell development (497)
 - cardiac muscle cell development (471)
 - cerebellar neuron development (...)
 - cerebral cortex GABAergic interneuron development (48)
 - chlamydospore formation (218)
 - chondrocyte development (279)
 - columnar/cuboidal epithelial cell development (818)
 - corticotropin hormone secreting cell development (...)

Ensembl genome browser 86 x Human Genome Overview ... x AmiGO 2: Drill-down Brow... x Semmelweis Egyetem Dokt... x

amigo.geneontology.org/amigo/dd_browse

No current user filters.

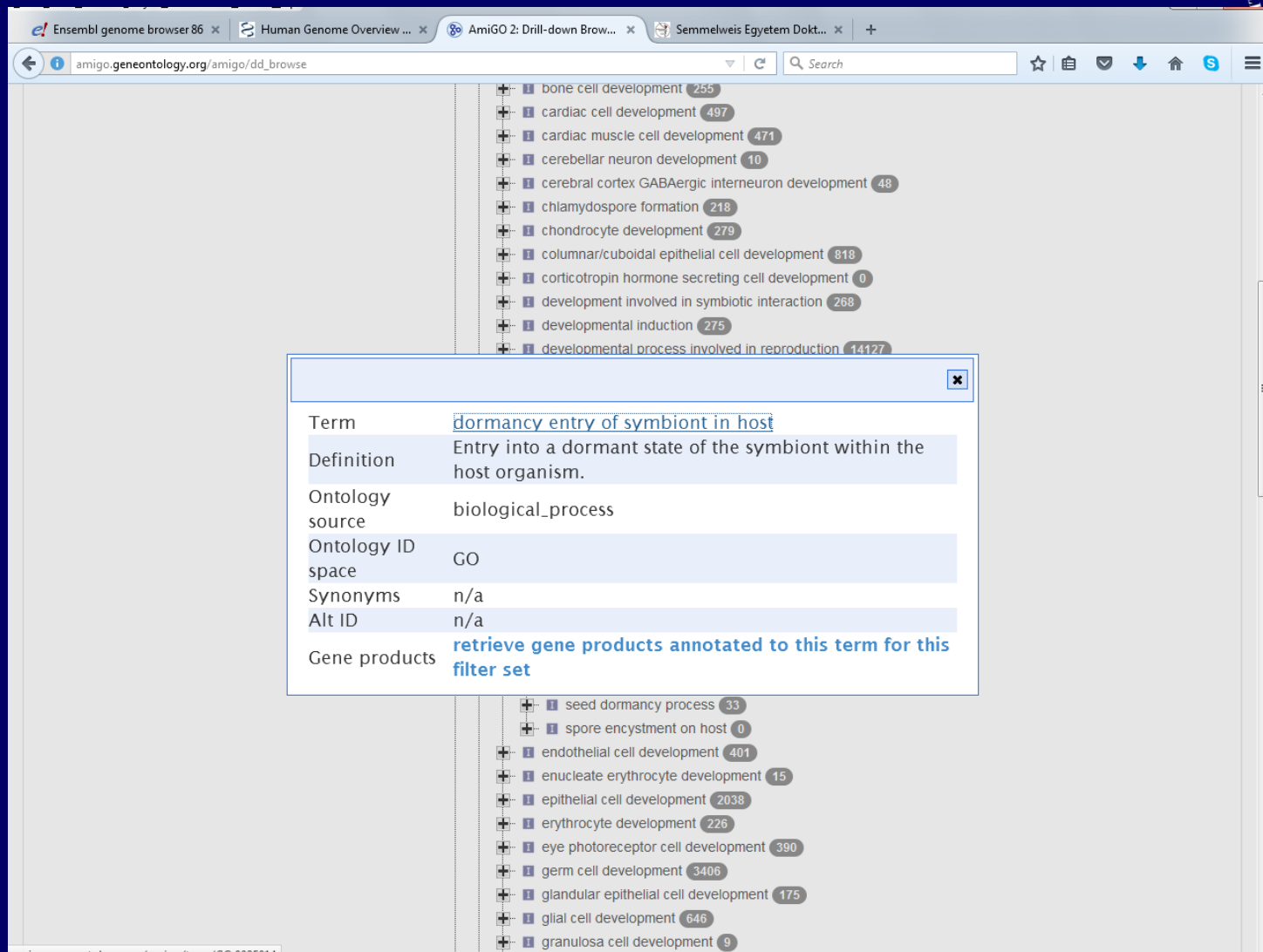
Your search is pinned to these filters

- document_category: bioentity

Organism

Type

- biological phase (244)
- biological regulation (239519)
- cell aggregation (341)
- cell development (23425)
- cell killing (1299)
- cellular component organization or biogenesis (140447)
- cellular process (510875)
- conidium development (1278)
- detoxification (4240)
- developmental process (84292)
 - Schwann cell development (202)
 - Sertoli cell development (99)
 - abscission (77)
 - acquisition of mycelium reproductive competence (0)
 - anatomical structure development (73952)
 - anatomical structure formation involved in morphogenesis (16739)
 - anatomical structure morphogenesis (35967)
 - astrocyte development (166)
 - auditory receptor cell development (184)
 - bone cell development (255)
 - cardiac cell development (497)
 - cardiac muscle cell development (471)
 - cerebellar neuron development (10)
 - cerebral cortex GABAergic interneuron development (48)
 - chlamydospore formation (218)
 - chondrocyte development (279)
 - columnar/cuboidal epithelial cell development (818)
 - corticotropin hormone secreting cell development (0)
 - development involved in symbiotic interaction (268)
 - developmental induction (275)
 - developmental process involved in reproduction (14127)
 - dormancy process (161)
 - acquisition of desiccation tolerance (2)
 - bud dormancy process (0)
 - dormancy entry of symbiont in host (2)
 - dormancy exit of symbiont in host (4)
 - dormancy maintenance of symbiont in host (0)



The screenshot shows the AmiGO 2 web interface for Gene Ontology (GO) terms. The browser tabs include 'Ensembl genome browser 86', 'Human Genome Overview ...', 'AmiGO 2: Drill-down Brow...', and 'Semmelweis Egyetem Dokt...'. The address bar shows 'amigo.geneontology.org/amigo/dd_browse'. A search bar is visible on the right. The main content area displays a list of GO terms with their counts in parentheses. A modal window is open, showing details for the term 'dormancy entry of symbiont in host'.

Term	dormancy entry of symbiont in host
Definition	Entry into a dormant state of the symbiont within the host organism.
Ontology source	biological_process
Ontology ID space	GO
Synonyms	n/a
Alt ID	n/a
Gene products	retrieve gene products annotated to this term for this filter set

The background list of GO terms includes:

- bone cell development (255)
- cardiac cell development (497)
- cardiac muscle cell development (471)
- cerebellar neuron development (10)
- cerebral cortex GABAergic interneuron development (48)
- chlamydospore formation (218)
- chondrocyte development (279)
- columnar/cuboidal epithelial cell development (818)
- corticotropin hormone secreting cell development (0)
- development involved in symbiotic interaction (268)
- developmental induction (275)
- developmental process involved in reproduction (14127)
- seed dormancy process (33)
- spore encystment on host (0)
- endothelial cell development (401)
- enucleate erythrocyte development (15)
- epithelial cell development (2038)
- erythrocyte development (226)
- eye photoreceptor cell development (390)
- germ cell development (3406)
- glandular epithelial cell development (175)
- glial cell development (646)
- granulosa cell development (9)

Ensembl genome browser 86 x Human Genome Overview ... x AmiGO 2: Drill-down Brow... x AmiGO 2: Search x Semmelweis Egyetem Dokt... x +

amigo.geneontology.org/amigo/search/bioentity?q=*&fq=regulates_closure:"GO:0085014"&sfq=document_category:"bi" Search

AmiGO 2 Home Search Browse Tools & Resources Help Feedback About AmiGO 1.8 Quick search Search

Information about **Genes and gene products** search

Filter results

Total gene product(s): 2

Free-text filter

User filters

+ regulates_closure: GO:0085014

Your search is pinned to these filters

- document_category: bioentity

Source

Organism

Type

PANTHER family

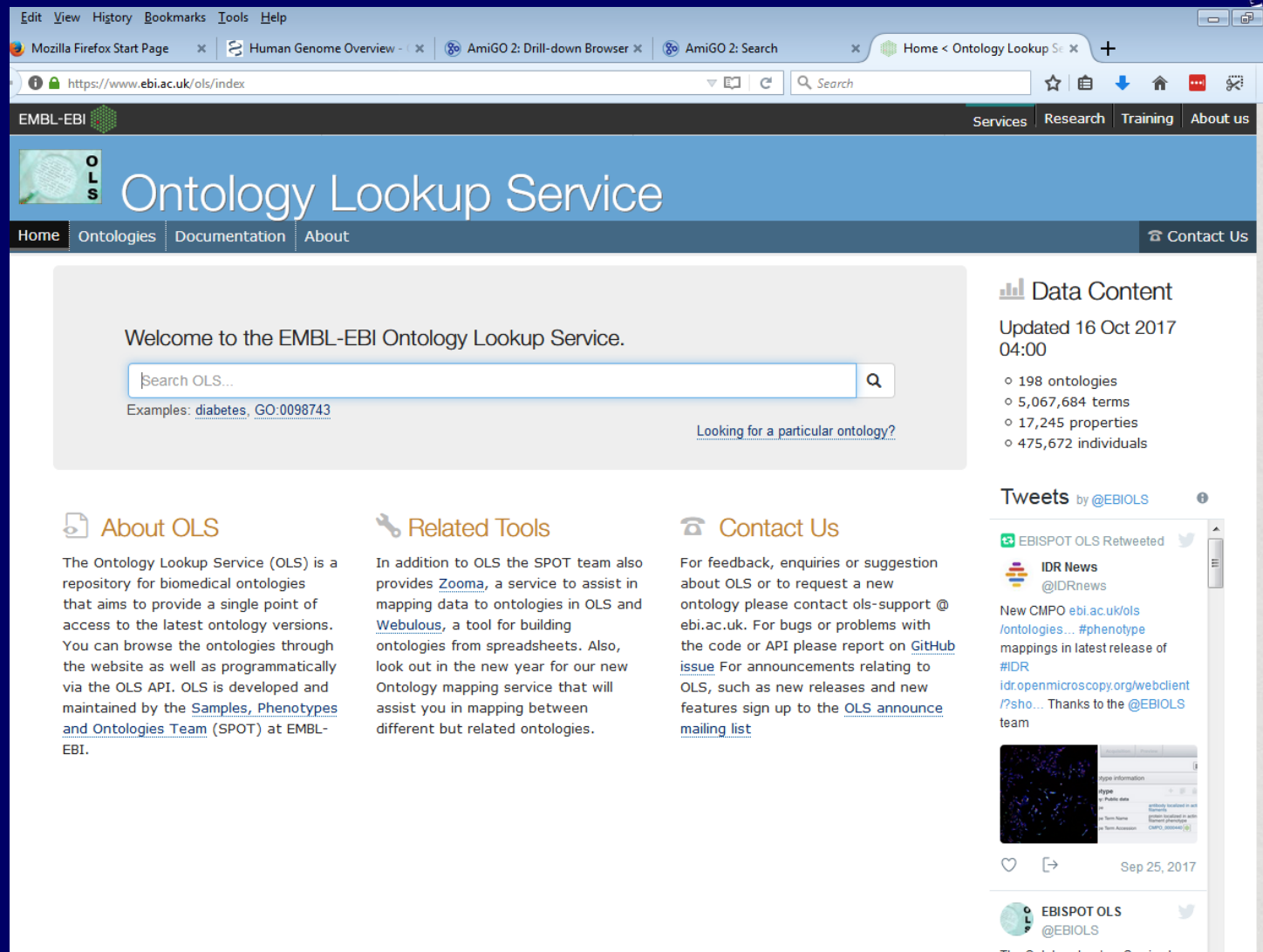
Direct annotation

Inferred annotation

Total gene product(s): 2; showing: 1-2
Results count 10

«First <Prev Next> Last» Custom DL (up to 100000) Bookmark

Gene/product	Gene/product name	Organism	PANTHER family	Type	Source	Synonyms
Rv2623	Universal stress protein Rv2623	Mycobacterium tuberculosis H37Rv		protein	UniProtKB	Y2623_MYCTU TB31.7
mce1A	MCE-family protein MCE1A	Mycobacterium tuberculosis H37Rv		protein	UniProtKB	Q79FZ9_MYCTU Rv0169 LH57_00945



Home < Ontology Lookup Service

Services Research Training About us

Ontology Lookup Service

Home Ontologies Documentation About Contact Us

Welcome to the EMBL-EBI Ontology Lookup Service.

Search OLS...

Examples: [diabetes](#), [GO:0098743](#)

[Looking for a particular ontology?](#)

About OLS

The Ontology Lookup Service (OLS) is a repository for biomedical ontologies that aims to provide a single point of access to the latest ontology versions. You can browse the ontologies through the website as well as programmatically via the OLS API. OLS is developed and maintained by the [Samples, Phenotypes and Ontologies Team](#) (SPOT) at EMBL-EBI.

Related Tools

In addition to OLS the SPOT team also provides [Zooma](#), a service to assist in mapping data to ontologies in OLS and [Webulous](#), a tool for building ontologies from spreadsheets. Also, look out in the new year for our new Ontology mapping service that will assist you in mapping between different but related ontologies.

Contact Us

For feedback, enquiries or suggestion about OLS or to request a new ontology please contact [ols-support@ebi.ac.uk](#). For bugs or problems with the code or API please report on [GitHub issue](#). For announcements relating to OLS, such as new releases and new features sign up to the [OLS announce mailing list](#).

Data Content

Updated 16 Oct 2017 04:00

- 198 ontologies
- 5,067,684 terms
- 17,245 properties
- 475,672 individuals

Tweets by @EBIOLS

EBISPOT OLS Retweeted

IDR News @IDRnews

New CMPO [ebi.ac.uk/ols/ontologies...](#) #phenotype mappings in latest release of #IDR

[idr.openmicroscopy.org/webclient/?sho...](#) Thanks to the @EBIOLS team

Sep 25, 2017

EBISPOT OLS @EBIOLS

Ensembl genome browse... x Human Genome Overvie... x AmiGO 2: Drill-down Bro... x AmiGO 2: Search x Semmelweis Egyetem Do... x Home < Ontology Looku... x +

www.ebi.ac.uk/ols/ontologies Search

This website uses cookies. By continuing to browse this site, you are agreeing to the use of our site cookies. To find out more, see our Terms of Use. OK

EMBL-EBI Services Research Training About us

Ontology Lookup Service

Home Ontologies Documentation About Contact Us

List of all ontologies in OLS

Enter an Ontology name Q

Ontology Name	Short name	Description	Loaded	Action
Anatomical Entity Ontology	AEO	AEO is an ontology of anatomical structures that expands CARO, the Common Anatomy Reference Ontology	Fri Jul 15 21:27:33 BST 2016	Search Terms Properties Individuals Download
Adverse Event Reporting Ontology	AERO	The Adverse Event Reporting Ontology (AERO) is an ontology aimed at supporting clinicians at the time of data entry, increasing quality and accuracy of reported adverse events	Tue May 17 01:33:25 BST 2016	Search Terms Properties Individuals Download
Agronomy Ontology	AGRO	Ontology of agronomic practices, agronomic techniques, and agronomic variables used in agronomic experiments	Wed Jul 27 06:51:47 BST 2016	Search Terms Properties Individuals Download
Ancestry Ontology	ancestro	The Ancestry Ontology (Ancestro) provides a systematic description of the ancestry concepts used in the NHGRI-EBI Catalog of published genome-wide association studies. It includes a list of countries, regions and major areas (essentially continents), as well as a fairly exhaustive list of country- or region-specific ancestral groups, uncategorised ancestral groups and population isolates.	Wed Mar 23 13:34:28 GMT 2016	Search Terms Properties Individuals Download
Ascomycete phenotype ontology	APO	A structured controlled vocabulary for the phenotypes of Ascomycete fungi	Fri Jul 15 21:36:32 BST 2016	Search Terms Properties Individuals Download

Az ENSEMBL analízis pipeline

- Az automatikus genomi annotáció a cél
- A kézi annotáció lassú és szubjektív
- A gépi gyors és konzisztens
- A legtöbb felhasználandó eszköz már kész
- Bemenő adatokbázisok: ENA, cDNS, UniProt, EST
- A pipe-line megfelelő sorrendben meghívja a programokat és az eredményt adatbázisba rendezi

A megvalósítás

- A pipeline moduláris szerkezetű – Runnable és RunnableDB
 - Ezek egy bizonyos feladatot végeznek
- Hierarchikus – RuleManager
 - Ez osztja ki a feladatokat és ellenőrzi az eredményt
- A számítások „computer farm”-on mennek
 - ~1000 node, közös adatbázist ér el

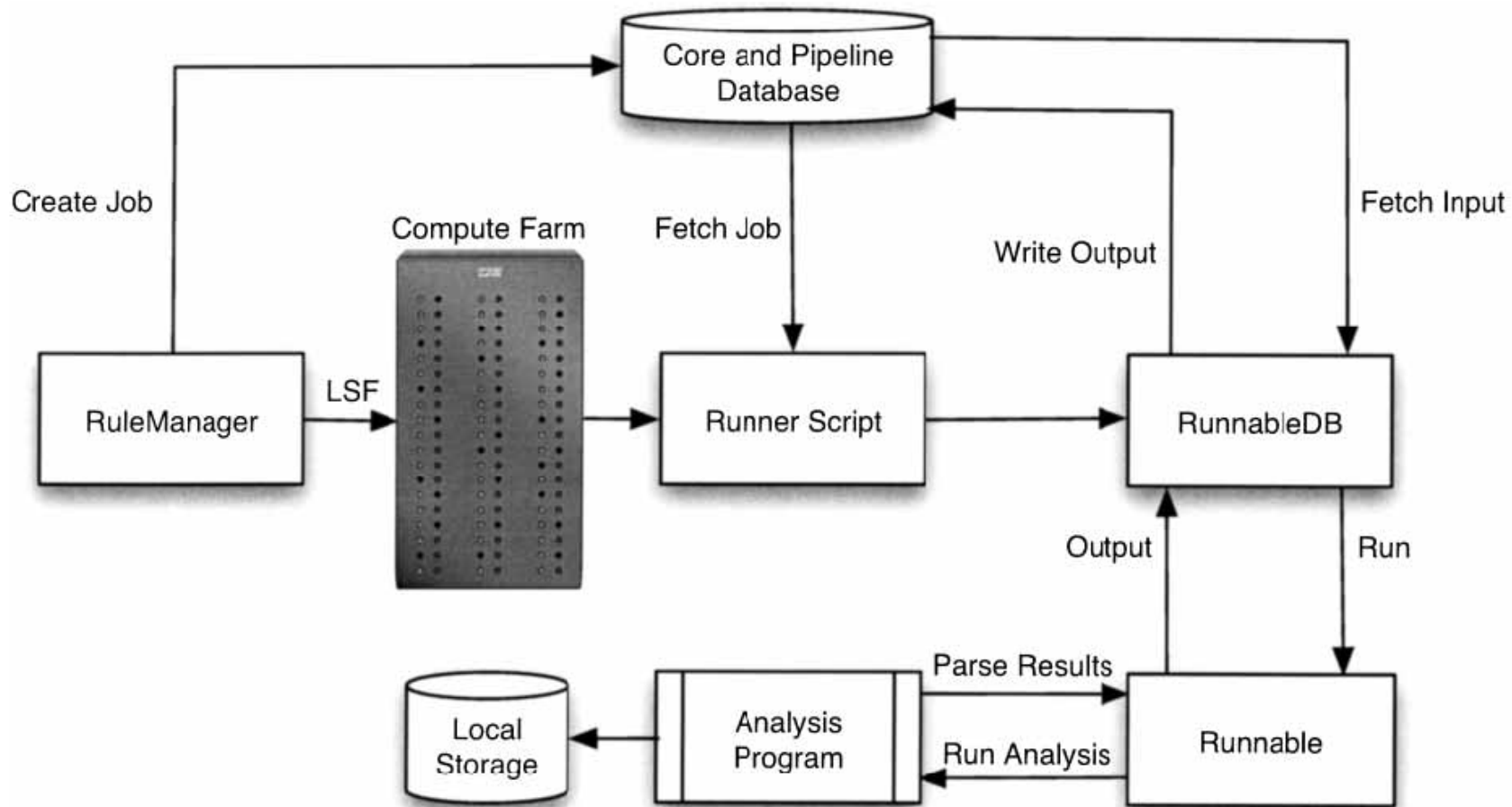


Figure 1 Ensembl pipeline system overview: The RuleManager uses LSF to submit analysis jobs to the compute farm. When an individual job starts executing on a remote node, the Runner script fetches the job information from the database and recreates the Job object. This in turn creates a RunnableDB and calls the appropriate methods (fetch_input, run, write_output, etc.) to run the analysis.

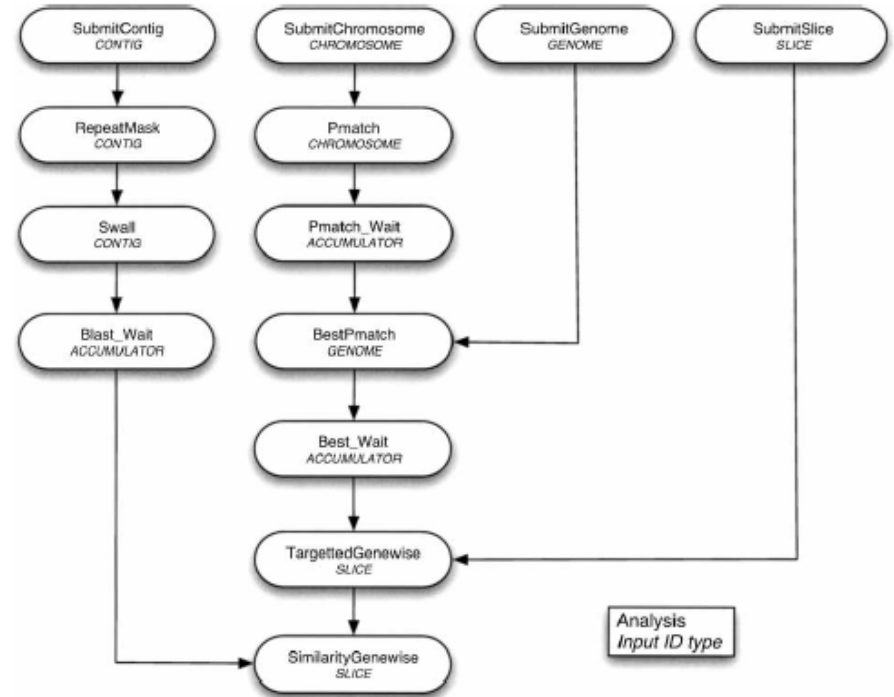
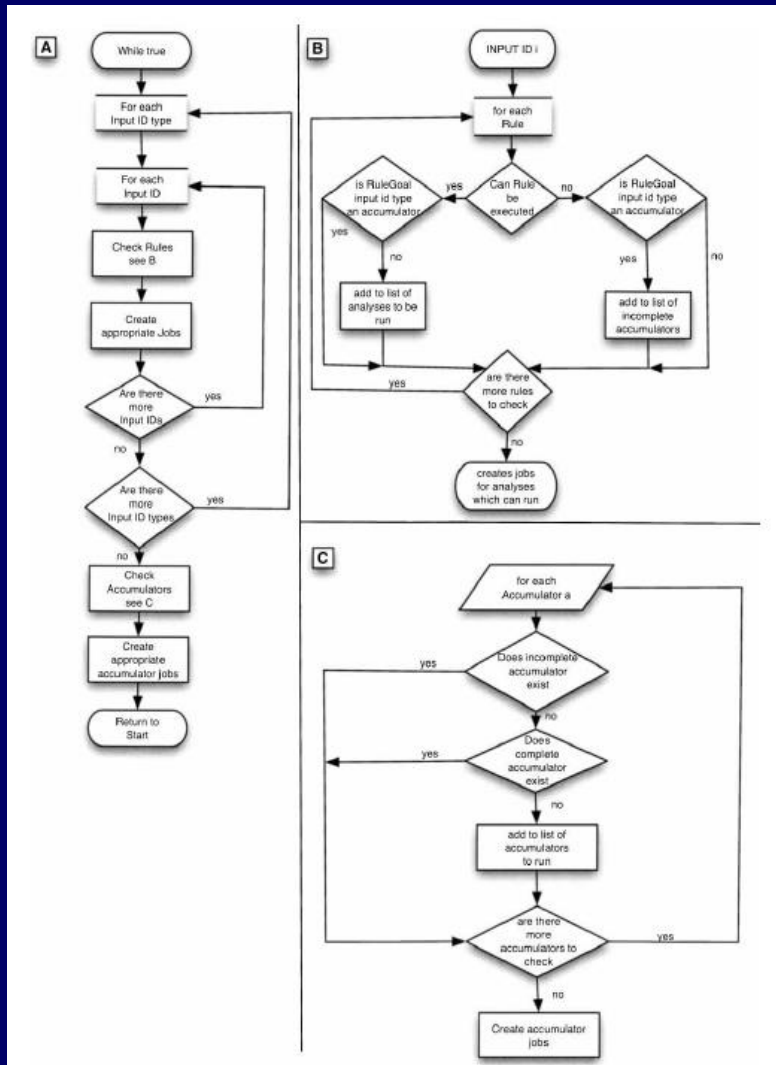


Figure 3 Pipeline control flow. This figure provides an example of the dependencies which can exist within the system. Analyses which have one dependency must maintain input id type, but analyses which have multiple dependencies can alter their input id type if an 'accumulator' analysis is used. (See the move from Swall to Similarity.genewise.)

Table 1. Principal Analyses of the 'Raw Compute' for the Human Genome With the Size of Input Sequence Used

Analysis	Input sequence size
CpG island prediction	chromosome
RepeatMasker	contig
Dust (low-complexity repeats)	chromosome
TRF (tandem repeats)	contig
Eponine (transcription start site prediction)	1-Mb slice
Genscan	contig
e-PCR (STS markers)	1-Mb slice
tRNAscan	contig
BLAST vs. Swall*	contig
BLAST vs. Unigene*	contig
BLAST vs. EMBL Vertebrate RNA*	contig

*BLAST analyses are only run on the peptides predicted by Genscan and not on the full genomic sequence. This is done to speed up the analysis.

Table 2. Principal Analyses of the Human Genome Protein Annotation

Analysis	Description	Input chunk size
Pfam	Interpro component	100 sequences
Prints	Interpro component	100 sequences
ScanProsite	Interpro component	whole protein data set
ProfileScan	Interpro component	100 sequences
Tmhmm	Transmembranes	100 sequences
ncoil	Coiled coils	100 sequences
sigp	Signal peptide	100 sequences
Seg	Low-complexity	whole protein data set

„Automatic Gene Annotation Sytem”

- Része az ENSEMBL pipeline-nak
- Célja:
 - Fehérje kódoló és RNS gének azonosítása
 - Exon-intron szerkezetük feltérképezése
 - Annotálása fehérje, cDNS és EST adatbázisok alapján
- A rendszer szabadon elérhető és használható

Az eljárás 'raw compute'

➤ Lépések:

- RepeatMasker – nagy számban előforduló szakaszok azonosítása
- Genescan – fehérje kódoló gének keresése
- tRNSscan – RNS gének keresése
- eponine – transkripció start-helyek keresése
- BLAST – fehérjék keresése homológia alapján
- A feldolgozás clone és contig szinten folyik
- Először gyors keresés – aztán pontosítás

Kombinált módszerek

- Az *Ab initio* génpredikció hasznos, de nem megbízható
- Genescan túl sok gént talál, viszont rövid exonokat hajlamos elveszíteni
- BLAST kereséssel meg kell támogatni az eredményt
- A BLAST viszont nem alkalmas exon-intron szerkezet megadására



Transzkript modellek

- Az adatbázis kiválasztása kulcsfontosságú
- A saját faj adatbázisa az elsődleges
- Más fajok adatbázisát is felhasználják a saját adat hiányában, de csak másodlagosan
- Ez áll a fehérje és cDNS adatokra is
- Így transzkript modelleket jósolunk először
- A két modell biztosan nem esik egybe a nem-transzlált régiók miatt (UTR)



Pmatch

- Gyorsan talál meg 100%-os egyezéseket
- A fehérje darabjainak egy szálon kell lenni
- A daraboknak jó sorrendben kell lenni
- A fehérje 25%-t legalább meg kell találni
- Egy fehérjéből a legjobb találatot megtartjuk
- Meg a továbbiakat is, ha csak 2% a különbség

genewise

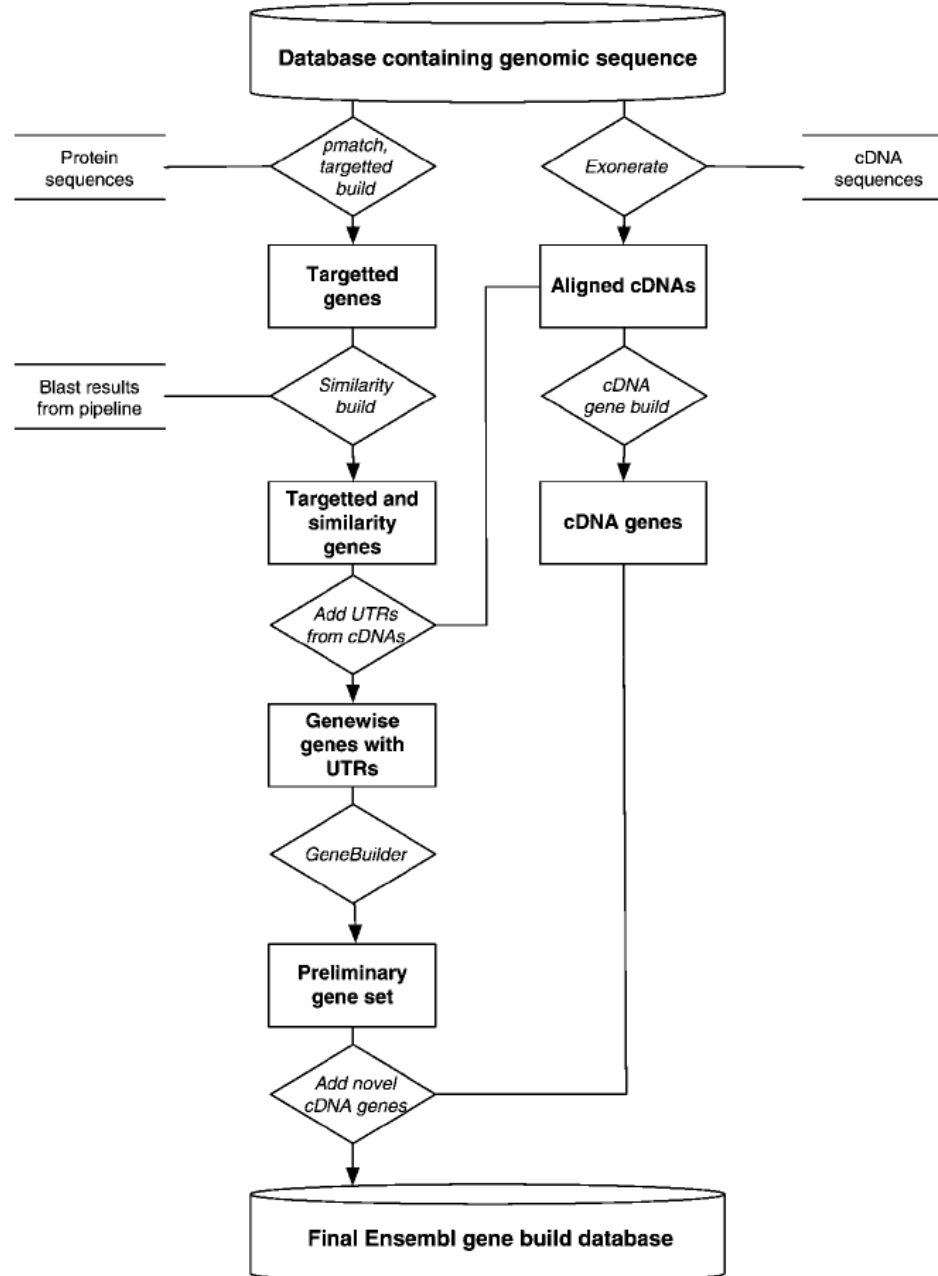
- A pmatch jelentősen lecsökkenti a vizsgálandó szegmensek hosszát
- Utána genewise – fehérje szintű illesztés splice-helyekkel és frame eltolással
- Viszont elég lassú
- *miniseq*: a megtalált nyers exonokat megtoldjuk 200 bázissal mindkét irányba
- Ez kezelhető méretűvé csökkenti a szekvenci hosszát

További finomítás

- Az eljárást megismételjük más fajokból származó fehérjékkel
- Csak az új találatokkal foglalkozunk
- A cDNS adatbázis saját fajhoz tartozó részét illesztjük a genomra (exonerate)
- A találatokat rangsoroljuk és kiválogatjuk
- A génszerkezetet kiegészítjük az UTR szakaszokkal

Végső modell (GeneBuilder)

1. Az egymással átfedő helyzetben lévő génmodelleket közös klaszterbe soroljuk
2. A közös exonnal rendelkező modelleket közös gén-klaszterekbe soroljuk
3. A nagyon nagy klasztereket megszüntjük és csak a legjobb 10-et tartjuk meg (ritka eset)
4. Ismét klaszterezzük a géneket, ha az előző lépés új klasztereket hozott volna létre



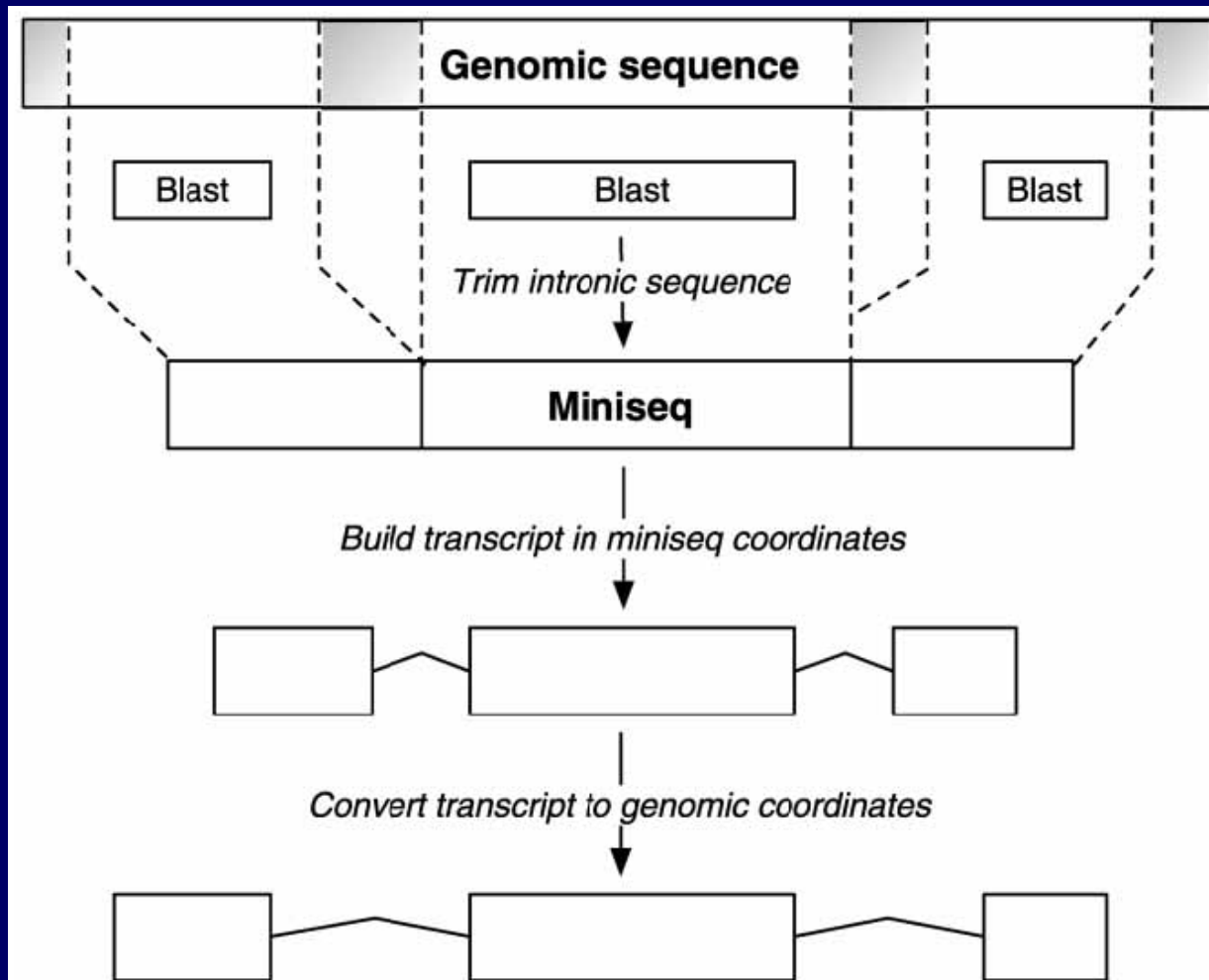
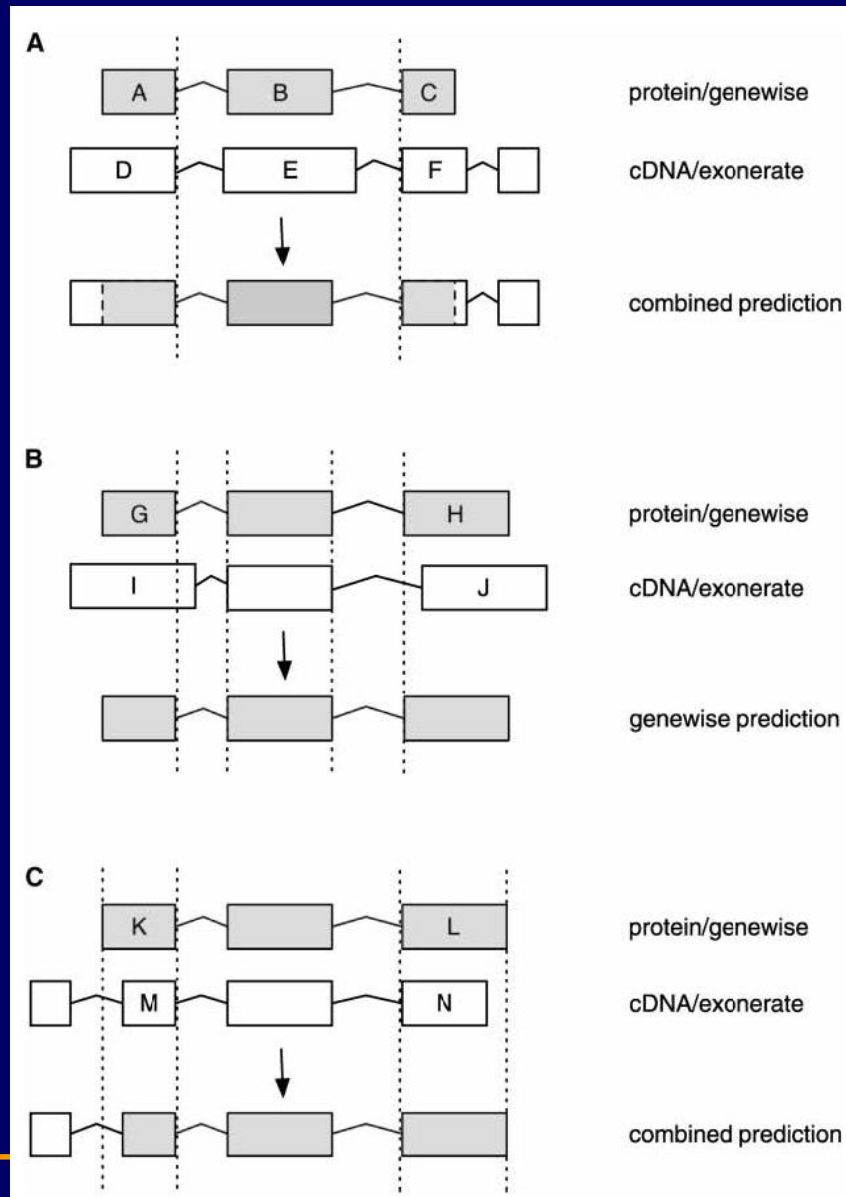
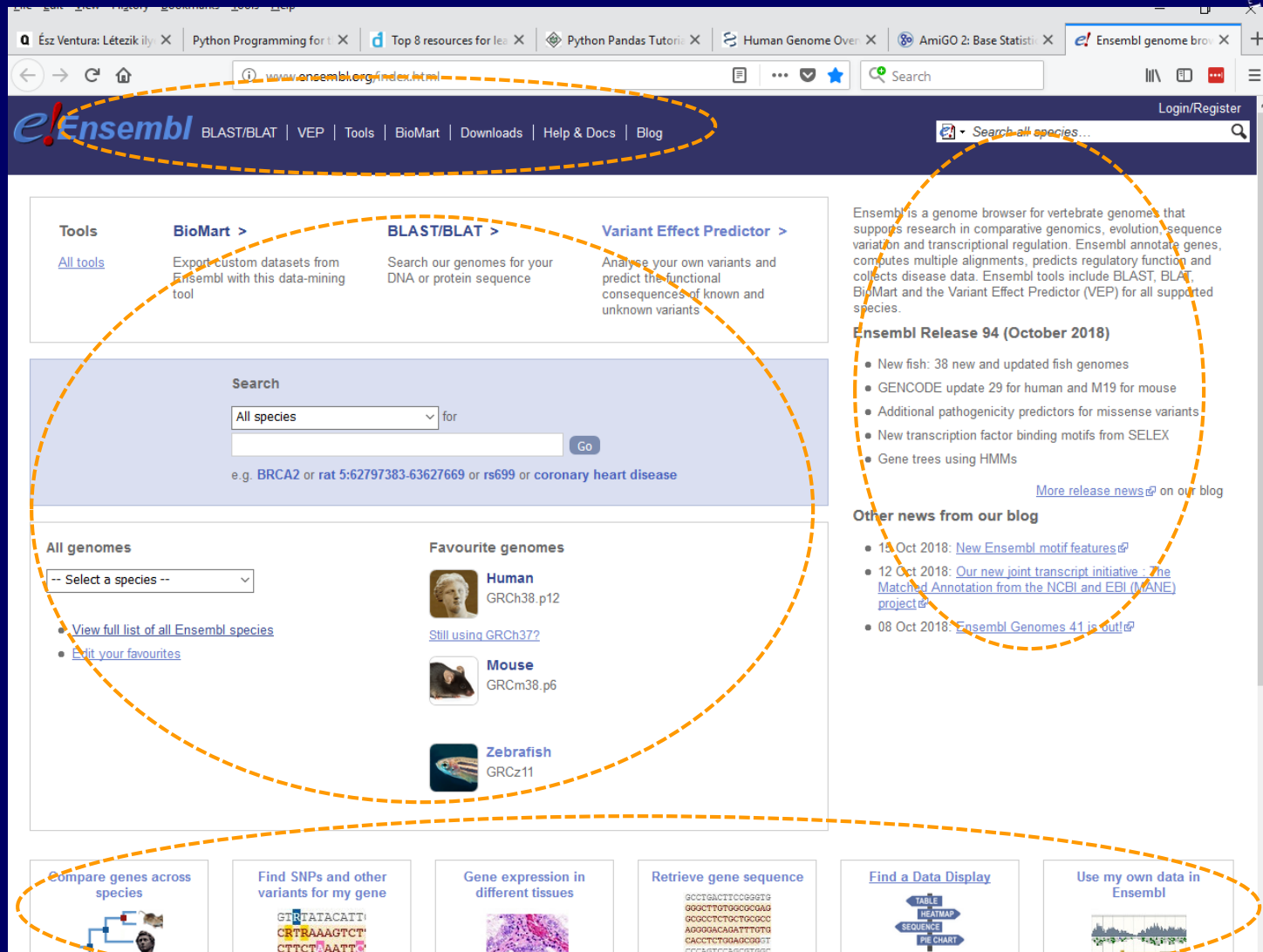


Figure 2 The Miniseq: We use a miniseq representation of genomic sequence in various stages of the gene build in order to reduce search space and increase processing speed. We BLAST a sequence of interest against a genomic region and pad the resulting hits with 200 bp. We then join the padded hits together to form a "mini genomic" sequence containing only exon sequence plus a small amount of intron sequence.



Az ENSEMBL felület

- Honlap: <http://www.ensembl.org/index.html>
- Közel 200 élőlény – többségében gerinces
- Ingyenesen elérhető, folyamatosan fejlesztik
- Közös Európai fenntartású kezdeményezés
- Több módon is elérhetők az adatok:
 - Web, BioMart, FTP
- Adatok feltöltése is lehetséges



The screenshot shows the Ensembl genome browser interface. Orange dashed circles highlight the following areas:

- Navigation Bar:** The top navigation bar containing links for BLAST/BLAT, VEP, Tools, BioMart, Downloads, Help & Docs, and Blog.
- Tools Section:** A row of tool links including All tools, BioMart, BLAST/BLAT, and Variant Effect Predictor.
- Search Bar:** A central search bar with a dropdown menu for 'All species' and a 'Go' button.
- Genomes Section:** A section titled 'All genomes' with a dropdown menu to select a species, and 'Favourite genomes' listing Human, Mouse, and Zebrafish.
- Footer Tools:** A row of tool links at the bottom including Compare genes across species, Find SNPs and other variants for my gene, Gene expression in different tissues, Retrieve gene sequence, Find a Data Display, and Use my own data in Ensembl.

Ensembl Release 94 (October 2018)

- New fish: 38 new and updated fish genomes
- GENCODE update 29 for human and M19 for mouse
- Additional pathogenicity predictors for missense variants
- New transcription factor binding motifs from SELEX
- Gene trees using HMMs

[More release news](#) on our blog

Other news from our blog

- 15 Oct 2018: [New Ensembl motif features](#)
- 12 Oct 2018: [Our new joint transcript initiative: The Matched Annotation from the NCBI and EBI \(MANE\) project](#)
- 08 Oct 2018: [Ensembl Genomes 41 is out!](#)

Ensembl genome browser interface showing the Species List page.

Browser tabs: Ensembl geno..., Human Genom..., AmiGO 2: Drill..., AmiGO 2: Search, Semmelweis Eg..., GO_0085014 <..., GO_0085014 <..., Species List

Address bar: www.ensembl.org/info/about/species.html

Ensembl logo and navigation links: BLAST/BLAT | BioMart | Tools | Downloads | Help & Documentation | Blog | Mirrors

Search bar: Search all species...

Left sidebar menu:

- In this section
 - Species List
 - Species Tree
 - Release Cycle
 - System Architecture
 - Mirror sites
 - Scientific Publications
 - Workshops
 - Geek for a Week
 - Contact Us
 - Collaboration
 - Mailing Lists
 - Job Vacancies
 - Legal
 - Browser Genome Release Agree
 - Disclaimer
 - Ensembl Image Reuse
 - Ensembl name and logo policy
 - Mirror policy
 - Privacy Statement
 - Software Licence
 - Code of Conduct
 - Acknowledgements
 - Credits for species images
 - Projects using Ensembl
 - Scientific Advisory Board

Search documentation... Go

Navigation: Help & Documentation | About the Ensembl Project | Species List

Find a Species

The main Ensembl site focuses on vertebrate genomes - [scroll down](#) for links to our sister sites covering invertebrates, plants, bacteria, etc.

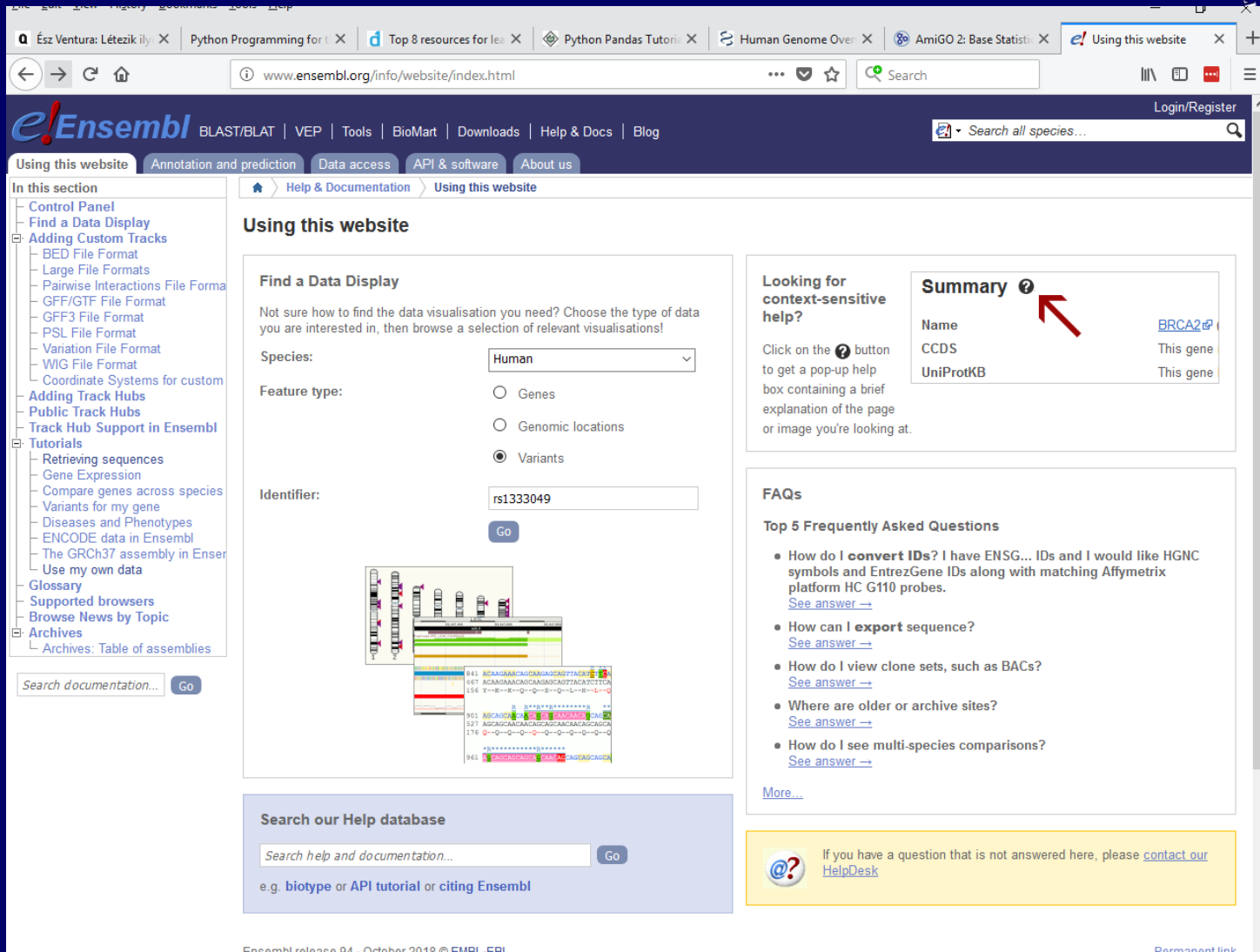
Species tree

[View the full Ensembl species tree.](#)

Ensembl Species

Note: to find out which species were in previous releases, please see the [table of assemblies](#)

Common name	Scientific name	Taxon ID	Ensembl Assembly	Accession	Variation database	Regulation database	Pre assembly
 Aardvark (Pre)	<i>Orycteropus afer afer</i>	1230840	-	-	-	-	OryAfe1
 Alpaca	<i>Vicugna pacos</i>	30538	vicPac1	-	-	-	-
 Amazon molly	<i>Poecilia formosa</i>	48698	Poecilia_formosa-5.1.2	GCA_000485575.1	-	-	-
 Anole lizard	<i>Anolis carolinensis</i>	28377	AnoCar2.0	GCA_000090745.1	-	-	-
 Armadillo	<i>Dasypus novemcinctus</i>	9361	Dasnov3.0	GCA_000208655.2	-	-	-
 Budgerigar (Pre)	<i>Melopsittacus undulatus</i>	13146	-	-	-	-	MelUnd6_3
 Bushbaby	<i>Otolemur gamettii</i>	30611	OtoGar3	GCA_000181295.3	-	-	-
 C.intestinalis	<i>Ciona intestinalis</i>	7719	KH	GCA_000224145.1	-	Y	-
 C.savignyi	<i>Ciona savignyi</i>	51511	CSAV 2.0	-	-	-	-



Ensembl BLAST/BLAT | VEP | Tools | BioMart | Downloads | Help & Docs | Blog

Search all species...

Using this website | Annotation and prediction | Data access | API & software | About us

In this section

- Control Panel
- Find a Data Display
- Adding Custom Tracks
 - BED File Format
 - Large File Formats
 - Pairwise Interactions File Format
 - GFF/GTF File Format
 - GFF3 File Format
 - PSL File Format
 - Variation File Format
 - WIG File Format
 - Coordinate Systems for custom
- Adding Track Hubs
- Public Track Hubs
- Track Hub Support in Ensembl
- Tutorials
 - Retrieving sequences
 - Gene Expression
 - Compare genes across species
 - Variants for my gene
 - Diseases and Phenotypes
 - ENCODE data in Ensembl
 - The GRCh37 assembly in Ensembl
 - Use my own data
- Glossary
- Supported browsers
- Browse News by Topic
- Archives
 - Archives: Table of assemblies

Search documentation... Go

Using this website

Find a Data Display

Not sure how to find the data visualisation you need? Choose the type of data you are interested in, then browse a selection of relevant visualisations!

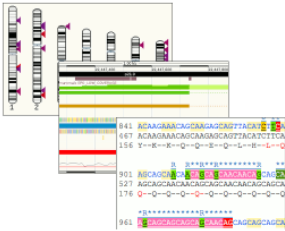
Species:

Feature type:

- ☐ Genes
- ☐ Genomic locations
- ☒ Variants

Identifier:

Go



Looking for context-sensitive help?

Click on the ? button to get a pop-up help box containing a brief explanation of the page or image you're looking at.

Summary ?

Name	BRCA2	This gene
CCDS		This gene
UniProtKB		This gene

FAQs

Top 5 Frequently Asked Questions

- How do I convert IDs? I have ENSG... IDs and I would like HGNC symbols and EntrezGene IDs along with matching Affymetrix platform HC G110 probes.
[See answer →](#)
- How can I export sequence?
[See answer →](#)
- How do I view clone sets, such as BACs?
[See answer →](#)
- Where are older or archive sites?
[See answer →](#)
- How do I see multi-species comparisons?
[See answer →](#)

[More...](#)

Search our Help database

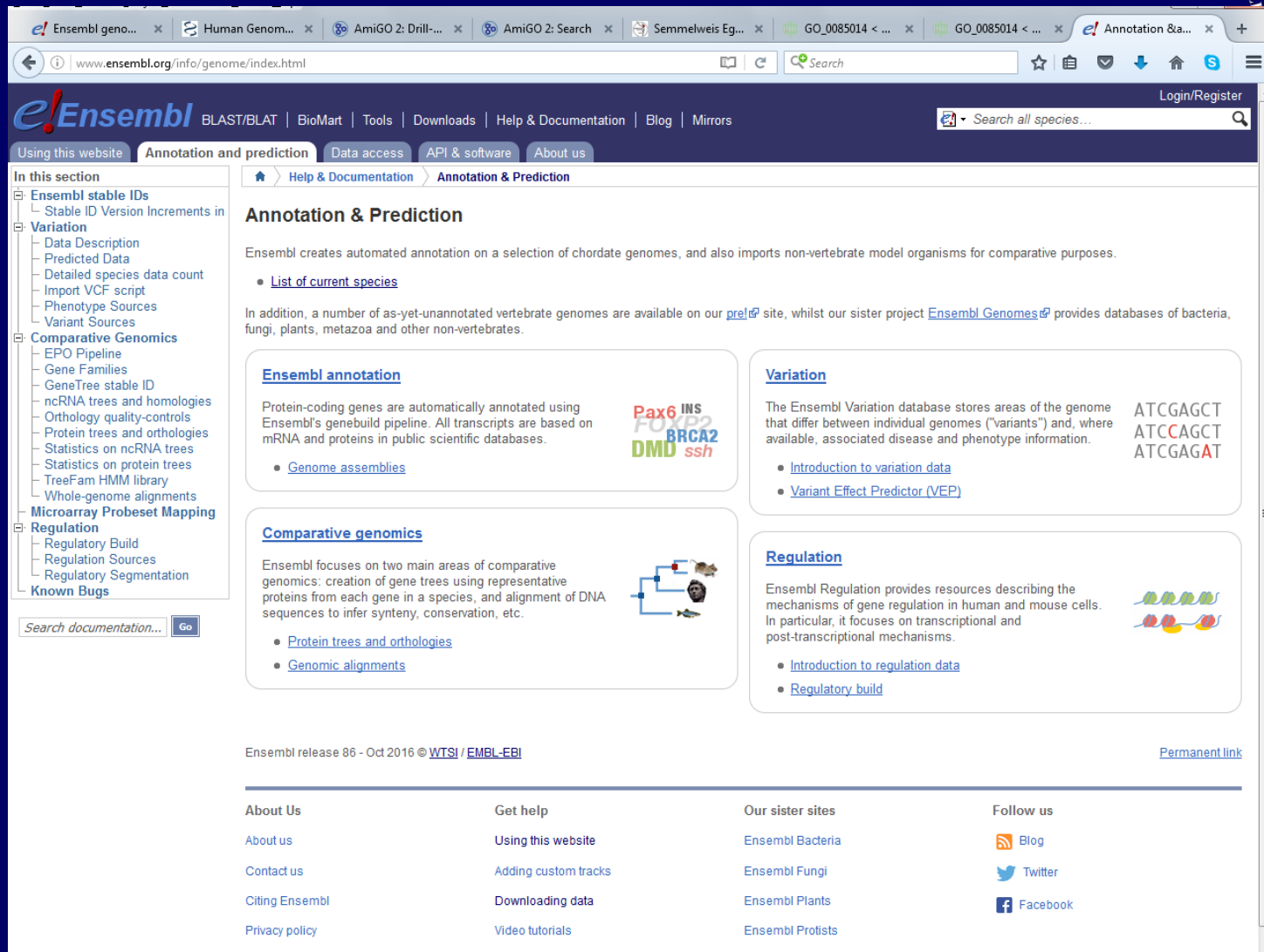
Search help and documentation... Go

e.g. biotype or API tutorial or citing Ensembl

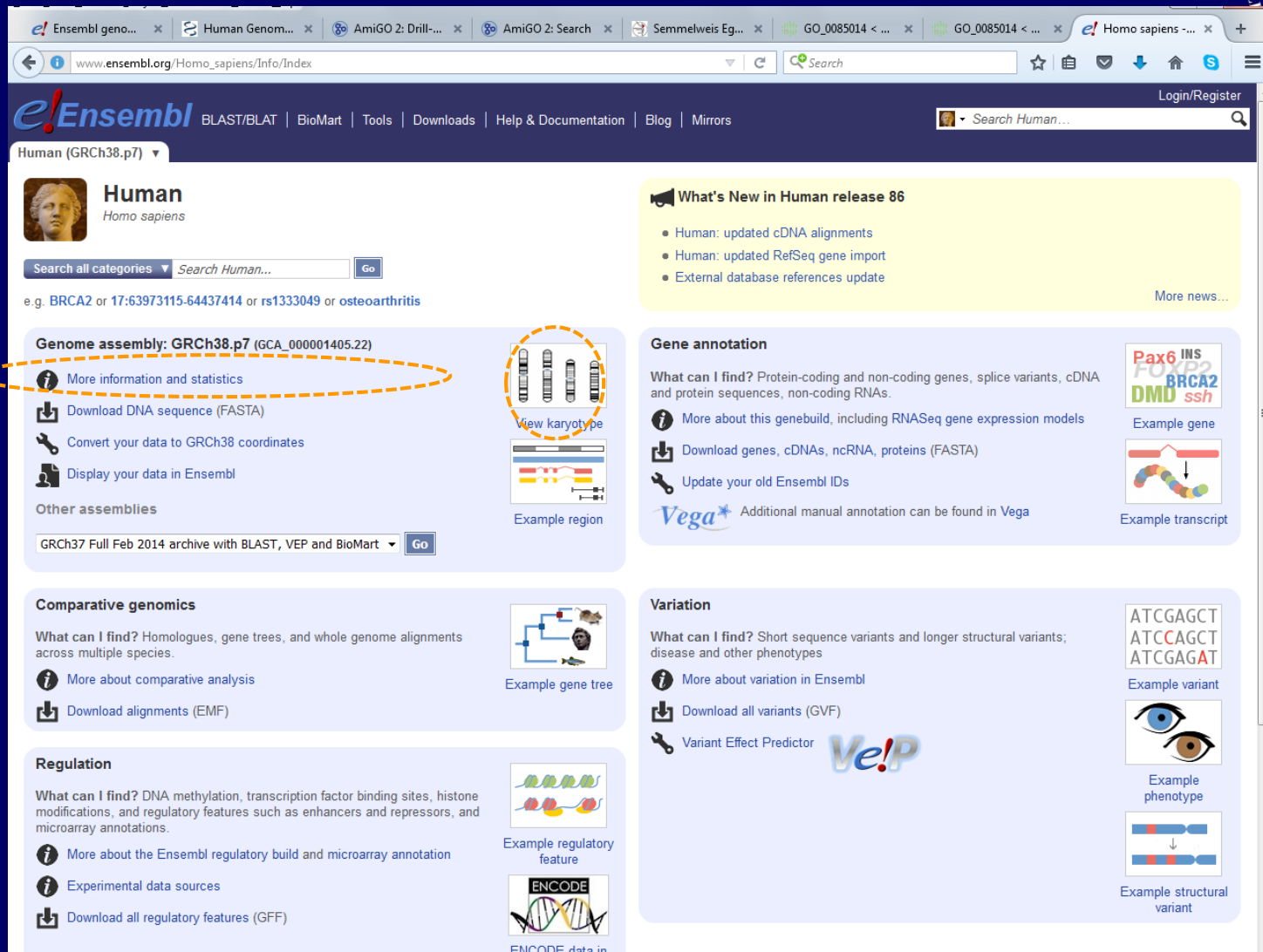
Ensembl release 94 - October 2018 © EMBL, EBI

[Permanent link](#)

If you have a question that is not answered here, please [contact our HelpDesk](#)



The screenshot shows the Ensembl genome browser interface. The browser tabs include 'Ensembl geno...', 'Human Genom...', 'AmiGO 2: Drill...', 'AmiGO 2: Search', 'Semmelweis Eg...', 'GO_0085014 < ...', and 'GO_0085014 < ...'. The address bar shows 'www.ensembl.org/info/genome/index.html'. The page title is 'Ensembl BLAST/BLAT | BioMart | Tools | Downloads | Help & Documentation | Blog | Mirrors'. The main navigation bar includes 'Using this website', 'Annotation and prediction', 'Data access', 'API & software', and 'About us'. The left sidebar lists 'In this section' with categories like 'Ensembl stable IDs', 'Variation', 'Comparative Genomics', 'Regulation', and 'Known Bugs'. The main content area is titled 'Annotation & Prediction' and describes Ensembl's automated annotation process. It includes links to 'List of current species', 'Ensembl annotation', 'Variation', 'Comparative genomics', and 'Regulation'. The 'Ensembl annotation' section mentions protein-coding genes and lists 'Genome assemblies'. The 'Variation' section mentions the Ensembl Variation database and lists 'Introduction to variation data' and 'Variant Effect Predictor (VEP)'. The 'Comparative genomics' section mentions gene trees and lists 'Protein trees and orthologies' and 'Genomic alignments'. The 'Regulation' section mentions gene regulation mechanisms and lists 'Introduction to regulation data' and 'Regulatory build'. The footer includes 'Ensembl release 86 - Oct 2016 © WTSI / EMBL-EBI', a 'Permanent link', and a table with links for 'About Us', 'Get help', 'Our sister sites', and 'Follow us'.



Ensembl genome browser interface for Human (GRCh38.p7).

Genome assembly: GRCh38.p7 (GCA_000001405.22)

- More information and statistics
- Download DNA sequence (FASTA)
- Convert your data to GRCh38 coordinates
- Display your data in Ensembl

Other assemblies

GRCh37 Full Feb 2014 archive with BLAST, VEP and BioMart

What's New in Human release 86

- Human: updated cDNA alignments
- Human: updated RefSeq gene import
- External database references update

Gene annotation

What can I find? Protein-coding and non-coding genes, splice variants, cDNA and protein sequences, non-coding RNAs.

- More about this genebuild, including RNASeq gene expression models
- Download genes, cDNAs, ncRNA, proteins (FASTA)
- Update your old Ensembl IDs

Vega Additional manual annotation can be found in Vega

Comparative genomics

What can I find? Homologues, gene trees, and whole genome alignments across multiple species.

- More about comparative analysis
- Download alignments (EMF)

Regulation

What can I find? DNA methylation, transcription factor binding sites, histone modifications, and regulatory features such as enhancers and repressors, and microarray annotations.

- More about the Ensembl regulatory build and microarray annotation
- Experimental data sources
- Download all regulatory features (GFF)

Variation

What can I find? Short sequence variants and longer structural variants; disease and other phenotypes

- More about variation in Ensembl
- Download all variants (GVF)
- Variant Effect Predictor

Ensembl genome browser interface showing the Whole genome view for Human (GRCh38.p7).

Location-based displays:

- Whole genome
- Chromosome summary
- Region overview
- Region in detail
- Comparative Genomics
 - Synteny
 - Alignments (image)
 - Alignments (text)
 - Region Comparison
- Genetic Variation
 - Resequencing
 - Strain table
 - Linkage Data
- Markers
- Other genome browsers
 - UCSC
 - NCBI
 - Ensembl GRCh37

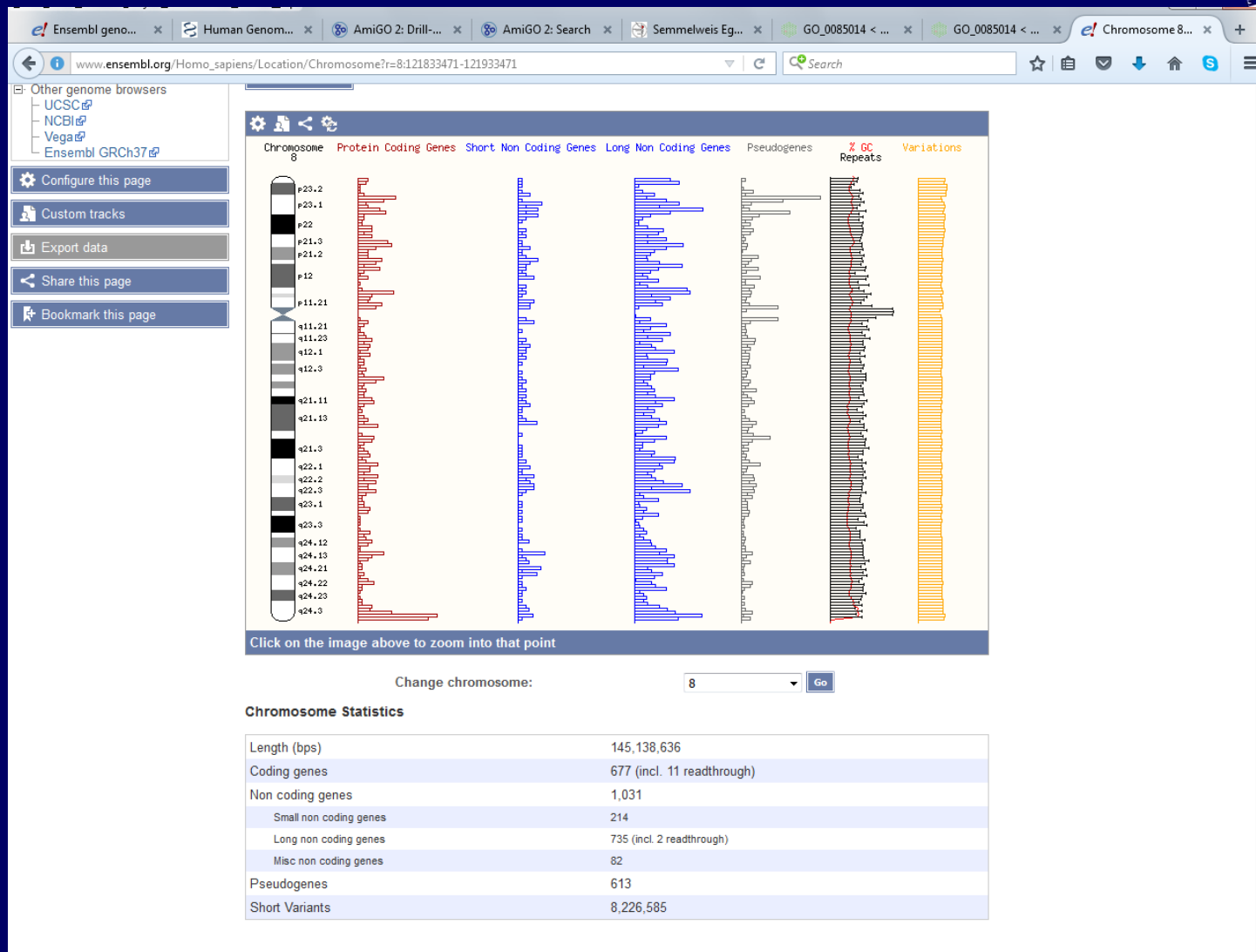
Whole genome

Click on the image above to jump to a chromosome, or click and drag to select a region

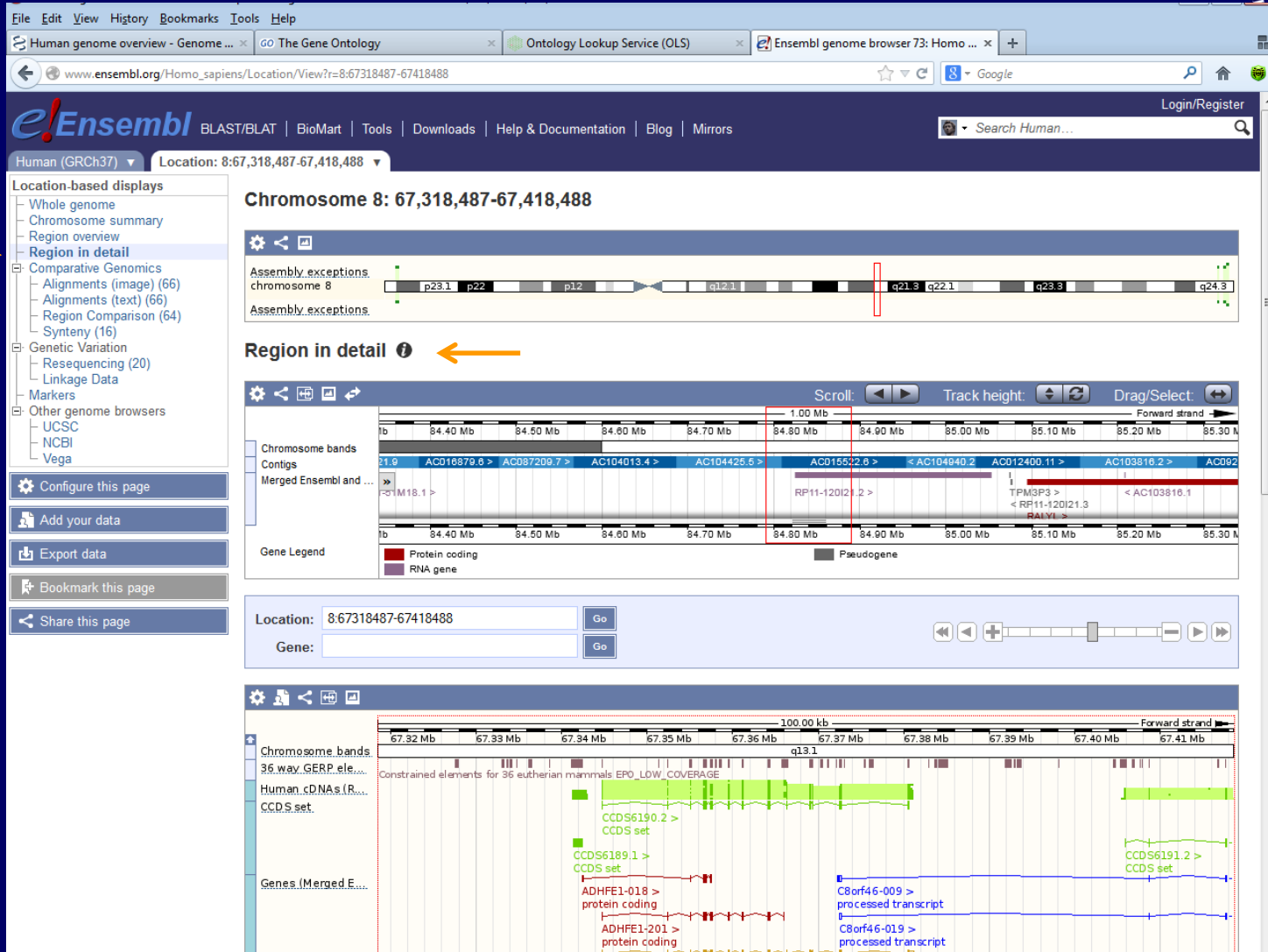
Summary

Assembly	GRCh38.p7 (Genome Reference Consortium Human Build 38), INSDC Assembly GCA_000001405.22 , Dec 2013
Database version	86.38
Base Pairs	3,547,762,741
Golden Path Length	3,096,649,726
Genebuild by	Ensembl
Genebuild method	Full genebuild
Genebuild started	Jan 2014
Genebuild released	Jul 2014
Genebuild last updated/patched	Jun 2016
Gencode version	GENCODE 25

Gene counts (Primary assembly)



9



The screenshot displays the Ensembl genome browser interface for Human (GRCh37) at location 8:67,318,487-67,418,488. The interface is divided into several sections:

- Left Sidebar:** Contains navigation options such as "Location-based displays", "Comparative Genomics", "Genetic Variation", "Markers", and "Other genome browsers". An orange arrow points to the "Region in detail" option under "Location-based displays".
- Main View Area:**
 - Chromosome 8: 67,318,487-67,418,488:** Shows a chromosome map with a red box highlighting a specific region.
 - Region in detail:** Provides a detailed view of the selected region, including tracks for "Chromosome bands", "Contigs", "Merged Ensembl and ...", and "Gene Legend". An orange arrow points to this section.
 - Gene Legend:** Indicates "Protein coding" (red) and "RNA gene" (purple).
- Bottom Section:** Displays a detailed view of the gene structure, including "Chromosome bands", "35 way GERP elements", "Human cDNAs (RefSeq)", "CCDS set", and "Genes (Merged Ensembl)". It shows the structure of the ADHFE1 gene, including exons and introns, and associated transcripts like C8orf46-009 and C8orf46-019.

www.ensembl.org/Help/View?id=140

e!Ensembl

Help

- Search
- Help topics
 - Frequently Asked Questions
 - Video Tutorials
 - Glossary
 - Contact HelpDesk

Region in detail

[Example page](#)

Region in detail allows you to browse genes, variations, sequence conservation, and other annotation along the genome. There are three main panels (or images), shown in Figure 1.

FIGURE 1 - REGION IN DETAIL

Human (GRCh37) Location: 12:25,357,725-25,483,878 Gene: KRAS

Chromosome 12: 25,357,725-25,483,878

Region of Interest

Haplotypes / Patches

Chromosome Image

1 Mb Overview Image * Scrollable in many browsers

Genes

Gene or region of interest

Add more data tracks

Genome

Zoomable Region Image

Transcripts (splice variants)

The screenshot shows the Ensembl genome browser interface. A 'saved configuration' box is highlighted. To its right, three buttons are shown: 'Reset configuration', 'Reset track order', and 'Add your data'. Arrows point from these buttons to a larger box labeled 'Go back to default settings'. Below this, another box labeled 'Input your own data' has an arrow pointing to the 'Add your data' button.

Add or change data tracks using the **configure this page** tool button at the left of the page, the cog wheel icons in the images, or click on the track name itself. Add data by navigating to the **Custom data** tab in the configuration window and clicking on **Add a custom track**.

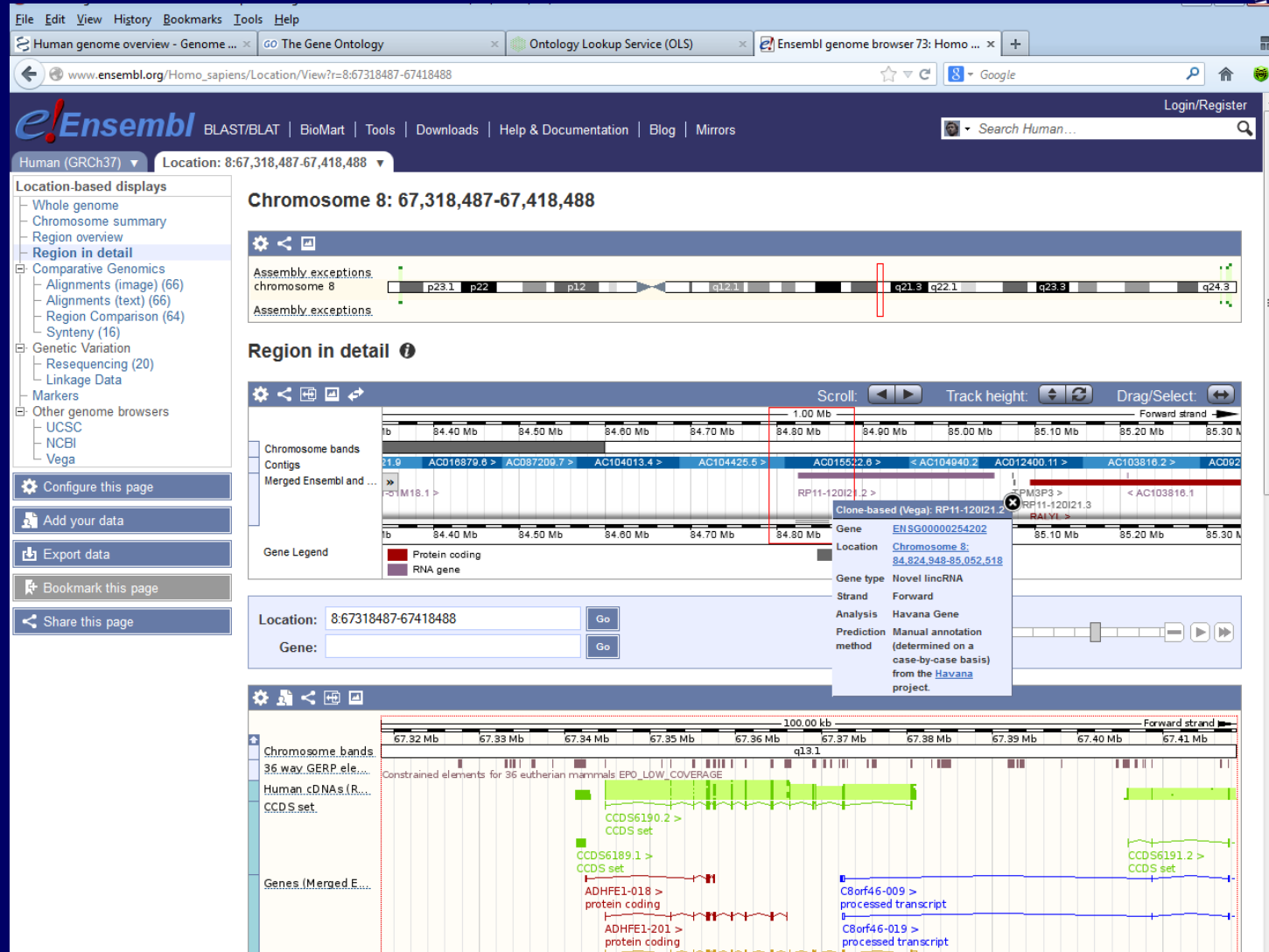
You can also [attach your own data](#) eg. BAM files by clicking the **Add your data** button (or **Manage your data** if you already have custom data added) and uploading or attaching a file, from the main view.

Some tracks can be displayed in different styles, however there is a limit on how much data can be displayed in certain styles. This [FAQ](#) explains the styles available and the amount of data that can be displayed.

Looking at the Region of the LCT Gene - The Region in Detail View

The screenshot shows the 'Region in detail' view for the LCT gene on Chromosome 2. The view displays various tracks including 'Chromosome bands', 'Genes (merged)', and 'Transcript (LCT-001)'. A play button icon is overlaid on the 'Transcript (LCT-001)' track.

POWERED BY
YouTube



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

Human (GRCh37) Location: 8:67,318,487-67,418,488

Chromosome 8: 67,318,487-67,418,488

Assembly exceptions chromosome 8

Region in detail

Location: 8:67318487-67418488

Gene: **ADHFE1-201**

Gene type: Novel lincRNA

Strand: Forward

Analysis: Havana Gene

Prediction: Manual annotation (determined on a case-by-case basis) from the Havana project.

Chromosome bands

36 way GERP element

Human cDNAs (RefSeq)

CCDS set

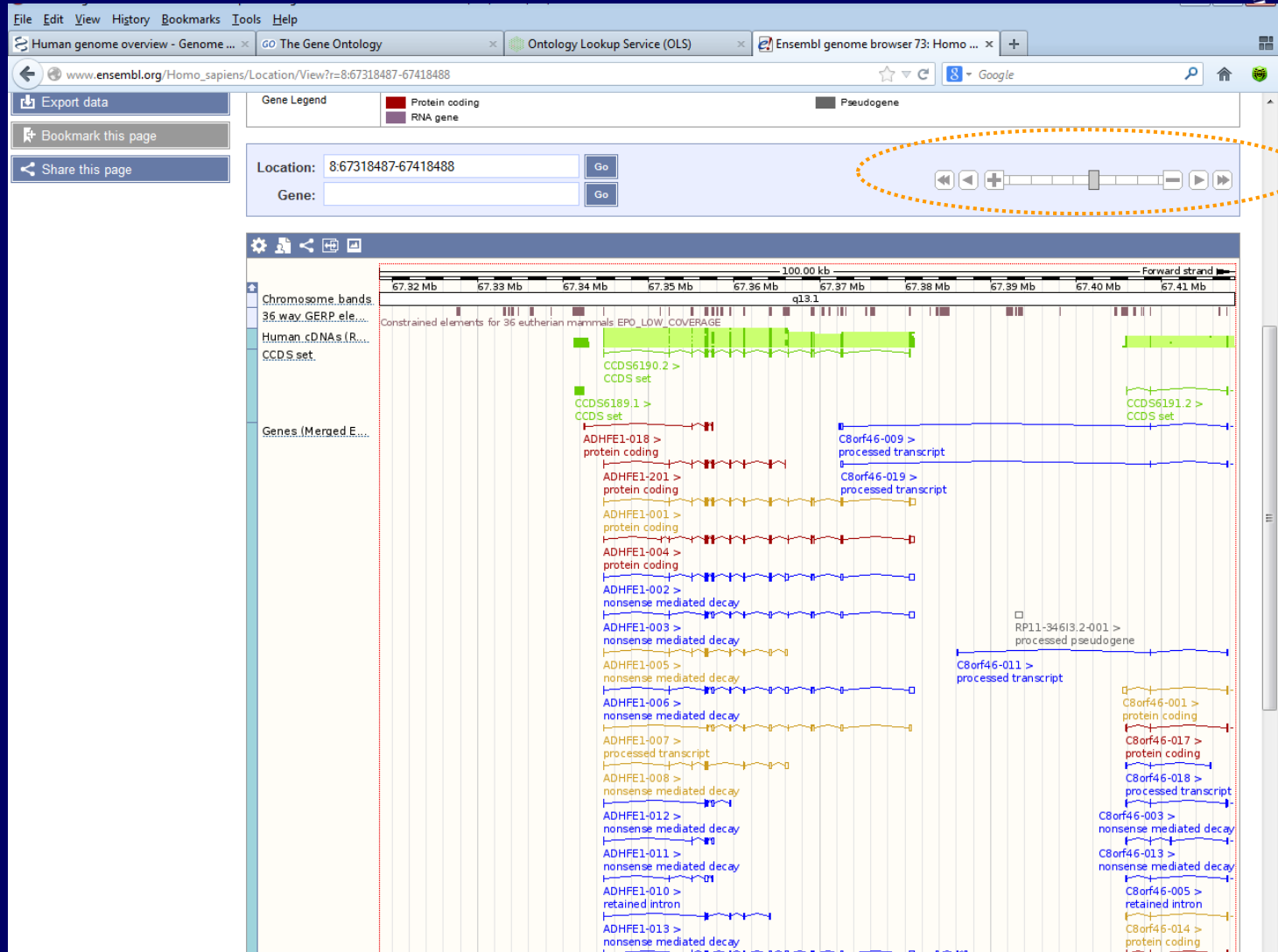
Genes (Merged Ensembl)

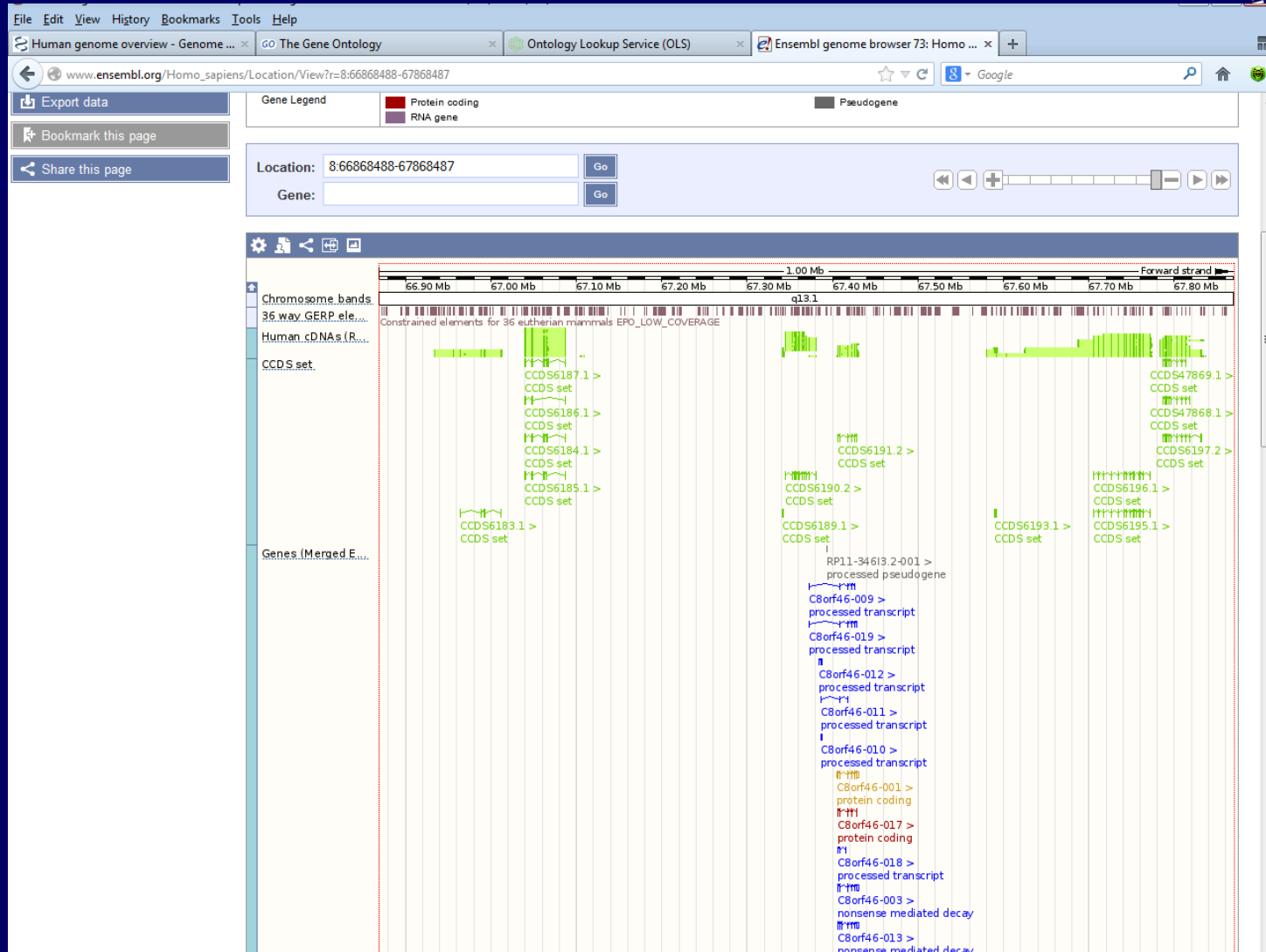
ADHFE1-018 > protein coding

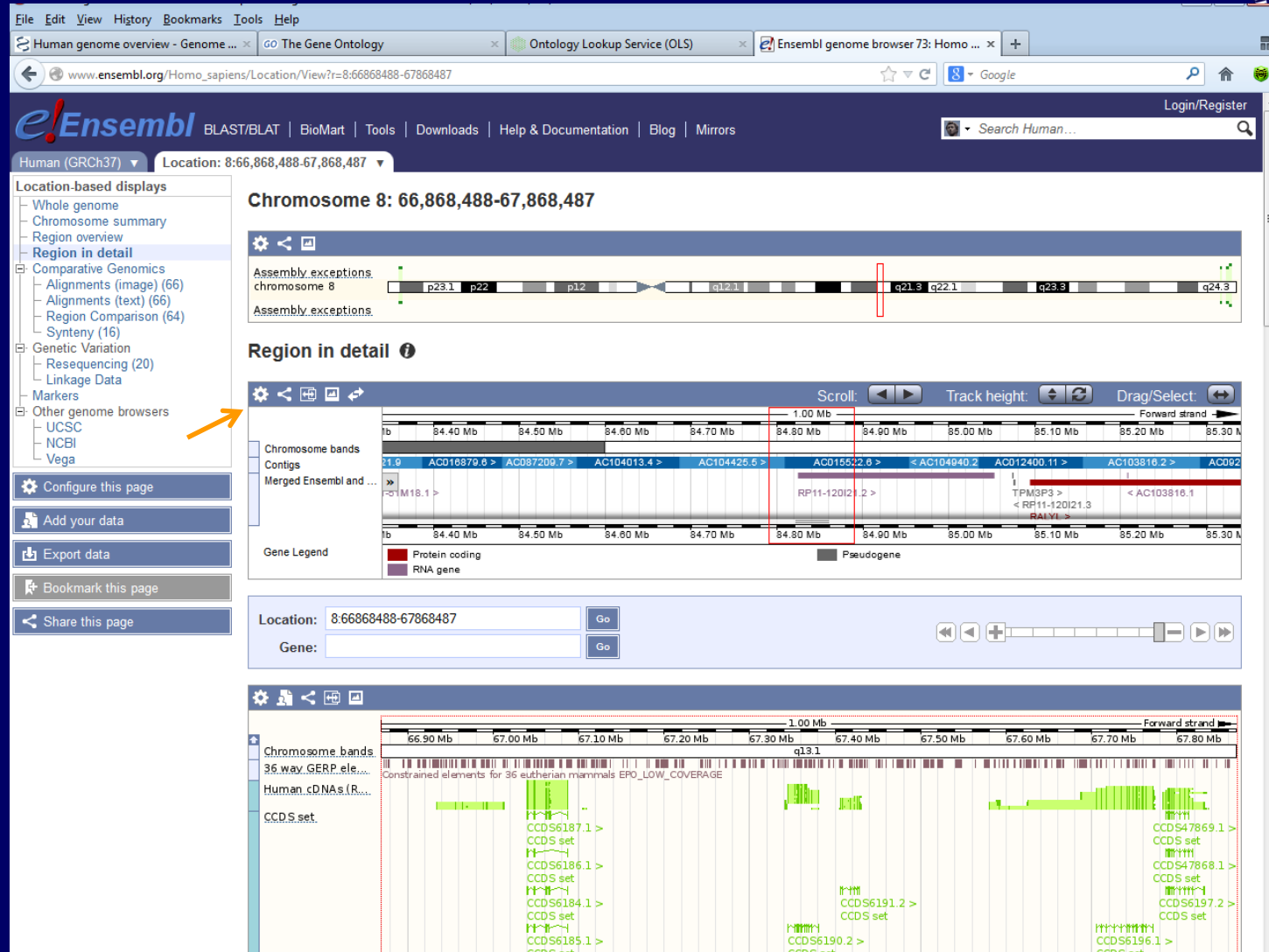
ADHFE1-201 > protein coding

C8orf46-009 > processed transcript

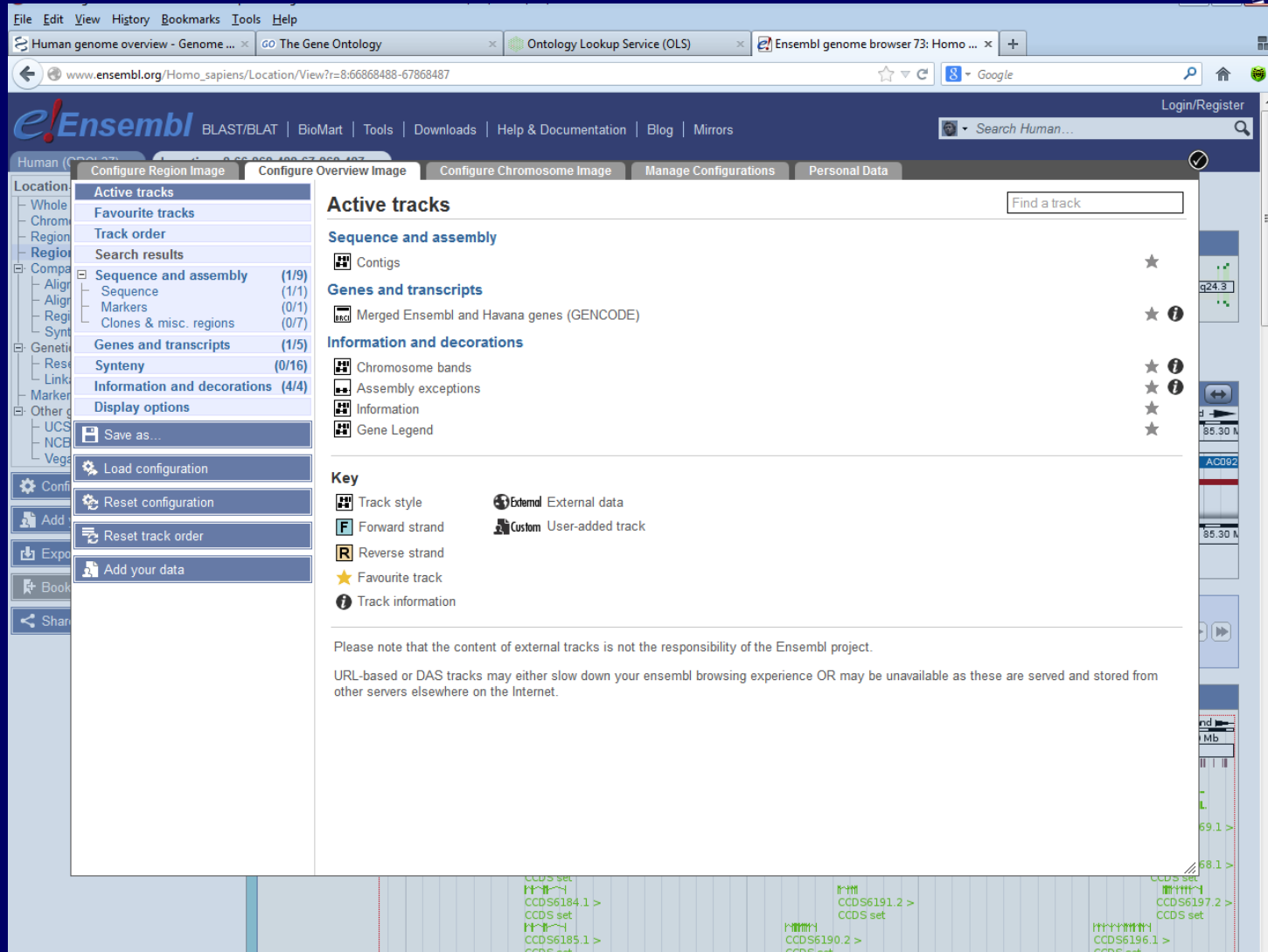
C8orf46-019 > processed transcript







The screenshot displays the Ensembl genome browser interface for Human (GRCh37) at location 8:66,868,488-67,868,487. The interface includes a top navigation bar with links to BLAST/BLAT, BioMart, Tools, Downloads, Help & Documentation, Blog, and Mirrors. A search bar is present on the right. The left sidebar contains a 'Location-based displays' menu with options like 'Whole genome', 'Chromosome summary', 'Region overview', 'Region in detail', 'Comparative Genomics', 'Genetic Variation', 'Markers', and 'Other genome browsers'. The main content area shows 'Chromosome 8: 66,868,488-67,868,487' with a genomic track. A red box highlights a specific region on the chromosome. The bottom section shows a detailed view of the '36 way GERP elements' and 'Human cDNAs (RefSeq)' for the region.



The screenshot shows the Ensembl genome browser interface with the 'Active tracks' configuration panel open. The browser window displays the URL www.ensembl.org/Homo_sapiens/Location/View?r=8:66868488-67868487. The 'Active tracks' panel lists various tracks under three categories: 'Sequence and assembly', 'Genes and transcripts', and 'Information and decorations'. The 'Sequence and assembly' category includes 'Contigs'. The 'Genes and transcripts' category includes 'Merged Ensembl and Havana genes (GENCODE)'. The 'Information and decorations' category includes 'Chromosome bands', 'Assembly exceptions', 'Information', and 'Gene Legend'. The 'Key' section defines symbols for 'Track style', 'Forward strand', 'Reverse strand', 'Favourite track', and 'Track information'. A note at the bottom states: 'Please note that the content of external tracks is not the responsibility of the Ensembl project. URL-based or DAS tracks may either slow down your ensembl browsing experience OR may be unavailable as these are served and stored from other servers elsewhere on the Internet.'

Active tracks

Sequence and assembly

- Contigs

Genes and transcripts

- Merged Ensembl and Havana genes (GENCODE)

Information and decorations

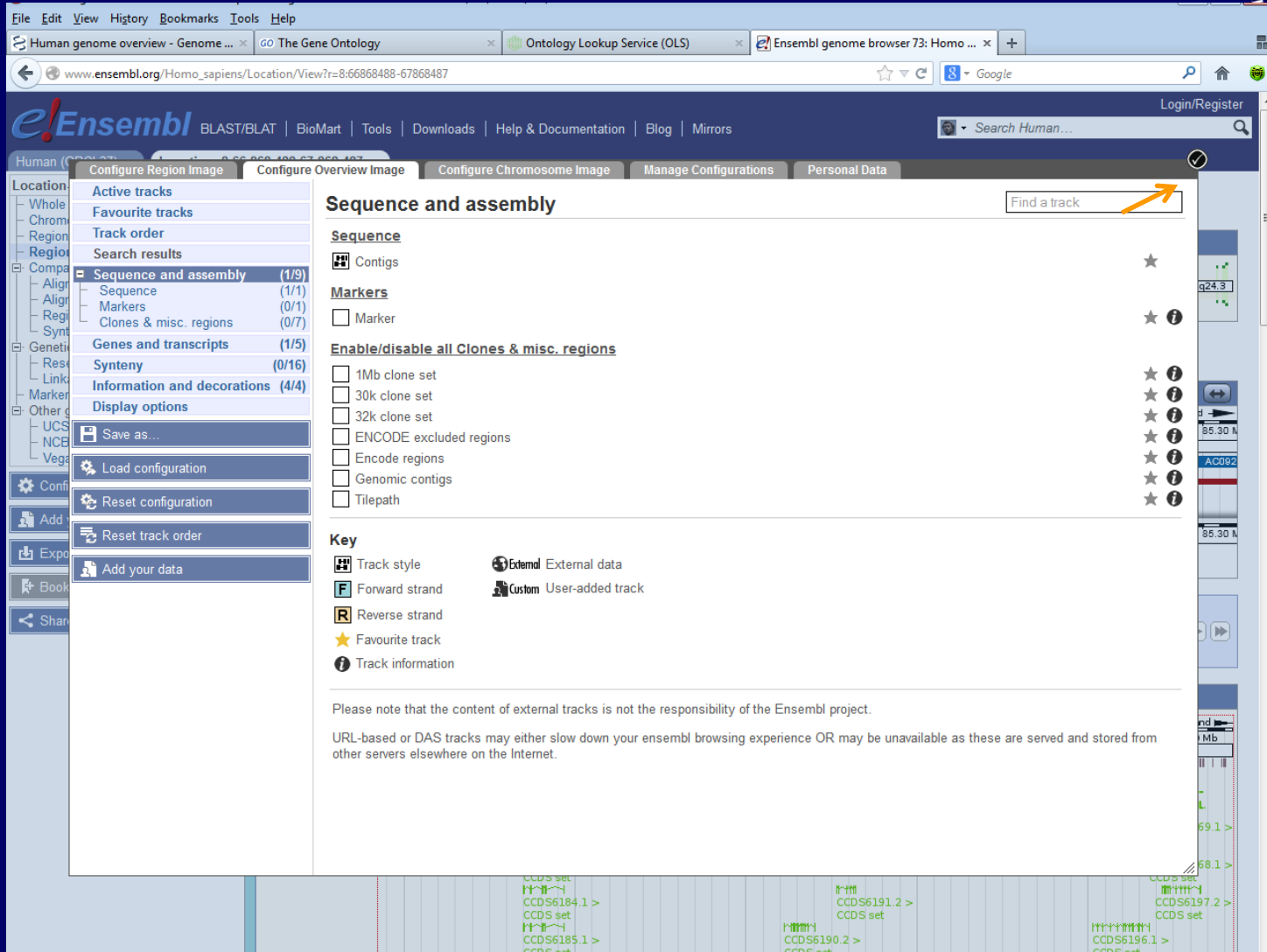
- Chromosome bands
- Assembly exceptions
- Information
- Gene Legend

Key

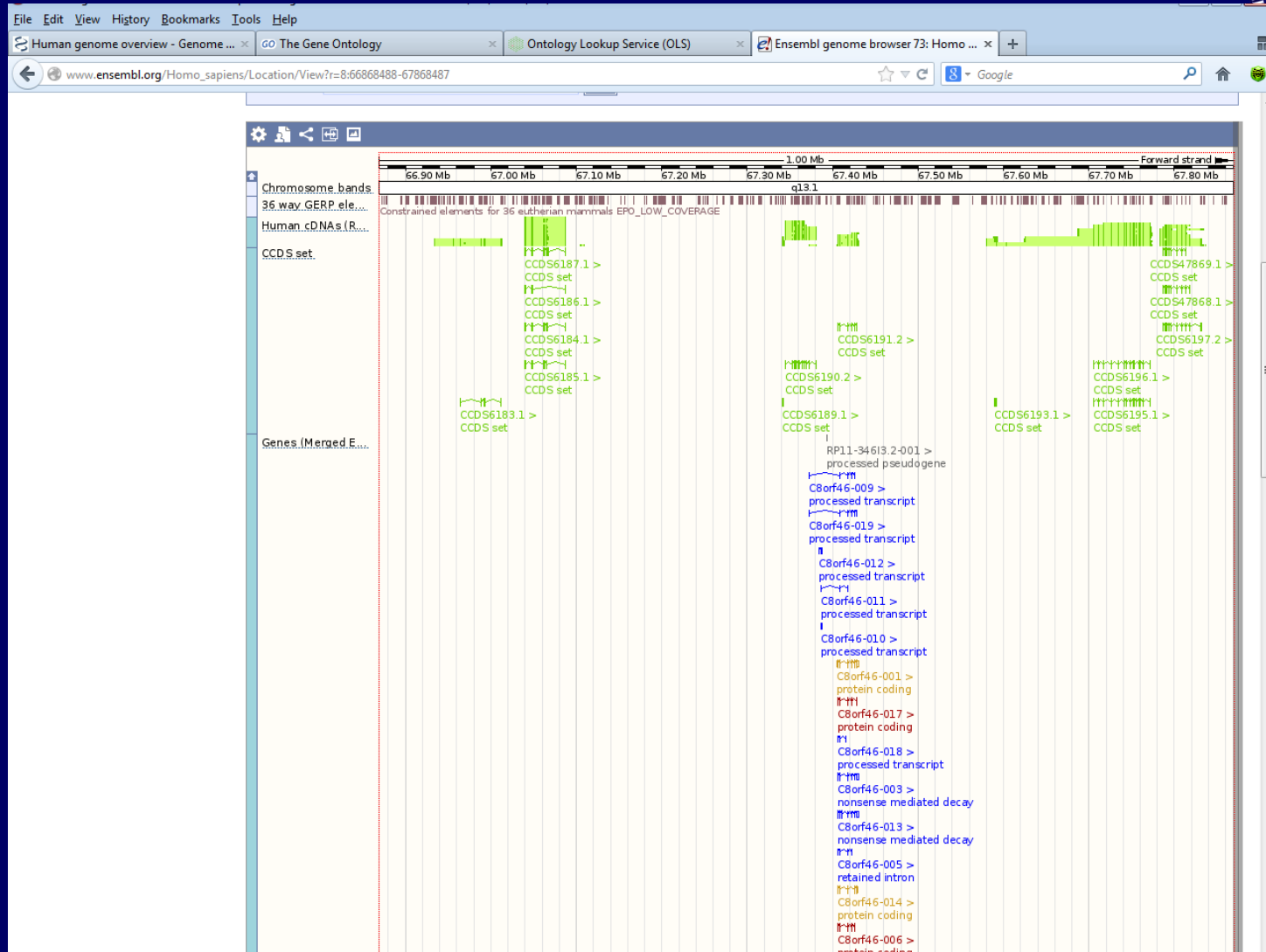
- Track style
- Forward strand
- Reverse strand
- Favourite track
- Track information

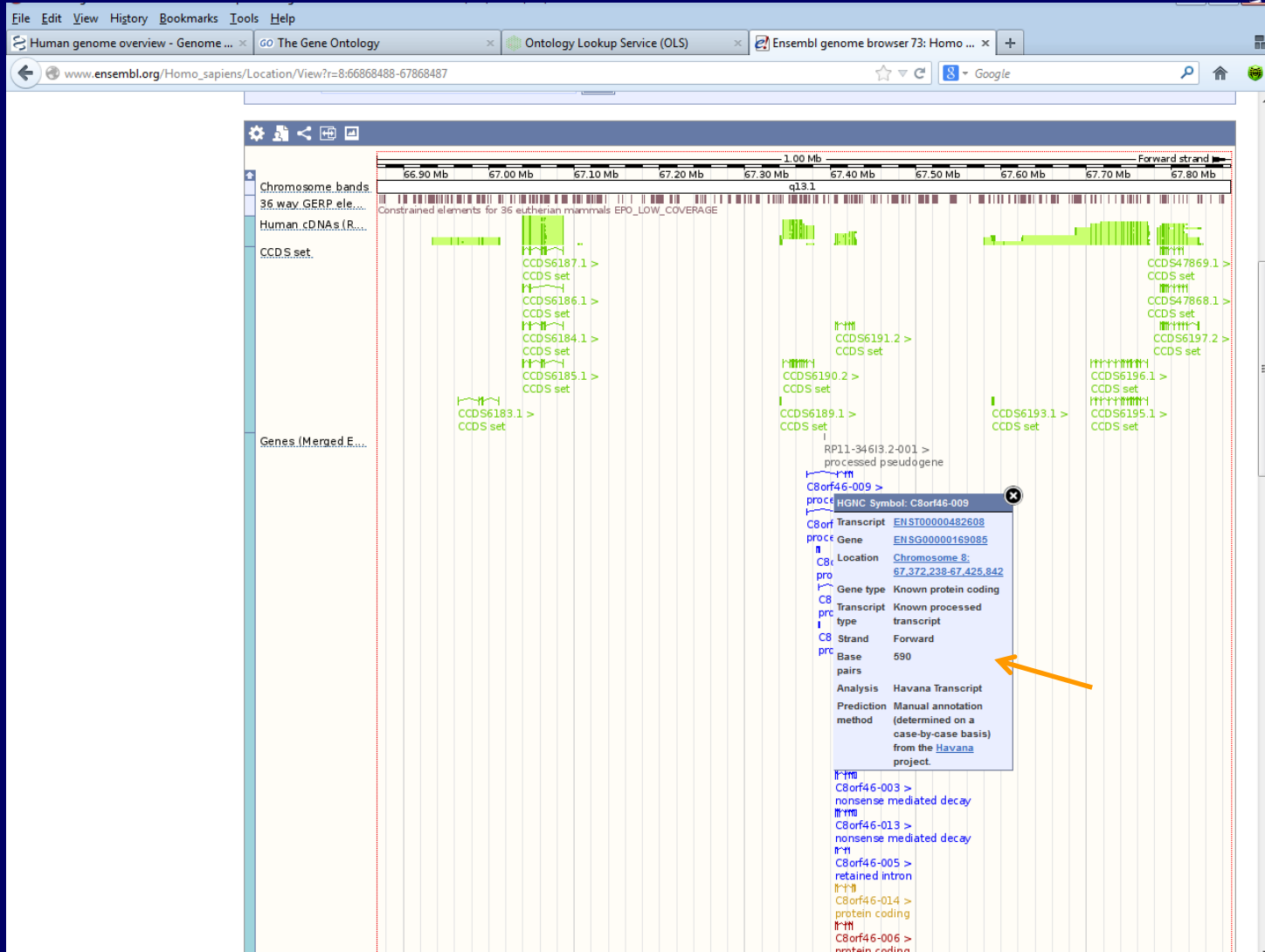
Please note that the content of external tracks is not the responsibility of the Ensembl project.

URL-based or DAS tracks may either slow down your ensembl browsing experience OR may be unavailable as these are served and stored from other servers elsewhere on the Internet.



The screenshot shows the Ensembl genome browser interface. The top navigation bar includes links for BLAST/BLAT, BioMart, Tools, Downloads, Help & Documentation, Blog, and Mirrors. A search bar is present with the text 'Search Human...'. The left sidebar shows a tree view of genomic features, with 'Sequence and assembly' selected. The main panel displays the 'Sequence and assembly' configuration options, including a 'Find a track' search bar, a list of tracks (Contigs, Markers, Clones & misc. regions, Genes and transcripts, Synteny, Information and decorations, Display options), and a 'Key' section for track styles (Track style, Forward strand, Reverse strand, Favourite track, Track information). A note at the bottom states: 'Please note that the content of external tracks is not the responsibility of the Ensembl project. URL-based or DAS tracks may either slow down your ensembl browsing experience OR may be unavailable as these are served and stored from other servers elsewhere on the Internet.'

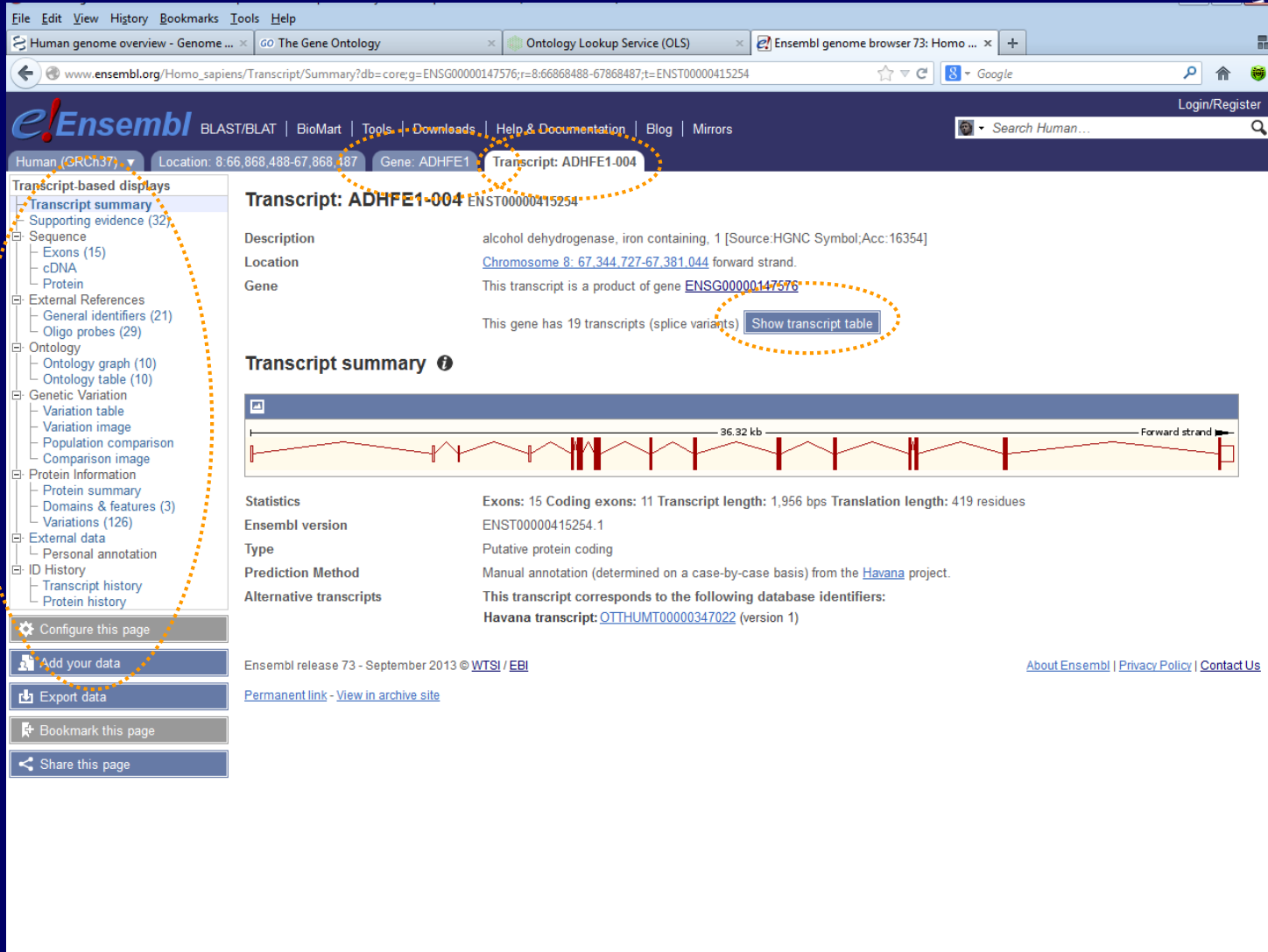




The screenshot shows the Ensembl genome browser interface. The top navigation bar includes links to 'Human genome overview', 'The Gene Ontology', 'Ontology Lookup Service (OLS)', and 'Ensembl genome browser 73: Homo sapiens'. The main track displays genomic data for a region on chromosome 8, from 66.90 Mb to 67.80 Mb. A pop-up window for the transcript 'C8orf46-009' is open, showing the following details:

Transcript	ENST00000482608
Gene	ENSG00000169085
Location	Chromosome 8: 67,372,238-67,425,842
Gene type	Known protein coding
Transcript type	Known processed transcript
Strand	Forward
Base pairs	590
Analysis	Havana Transcript
Prediction method	Manual annotation (determined on a case-by-case basis) from the Havana project.

Below the pop-up window, a list of other transcripts is visible, including C8orf46-003, C8orf46-013, C8orf46-005, C8orf46-014, and C8orf46-006.



The screenshot displays the Ensembl genome browser interface for the transcript ADHFE1-004. The browser window shows the URL: www.ensembl.org/Homo_sapiens/Transcript/Summary?db=core;g=ENSG00000147576;r=8:66868488-67868487;t=ENST00000415254. The left sidebar contains a tree view of transcript-based displays, including Transcript summary, Supporting evidence (32), Sequence (Exons (15), cDNA, Protein), External References, General identifiers (21), Oligo probes (29), Ontology, Genetic Variation, Protein Information, External data, ID History, Transcript history, and Protein history. The main content area shows the transcript details for ADHFE1-004, including its description, location, gene, and a transcript summary. The transcript summary includes a genomic track showing the transcript structure with exons and introns, and a table of statistics.

Transcript: ADHFE1-004 ENST00000415254

Description alcohol dehydrogenase, iron containing, 1 [Source:HGNC Symbol;Acc:16354]

Location [Chromosome 8: 67,344,727-67,381,044](#) forward strand.

Gene This transcript is a product of gene [ENSG00000147576](#). This gene has 19 transcripts (splice variants) [Show transcript table](#)

Transcript summary ⓘ

Statistics Exons: 15 Coding exons: 11 Transcript length: 1,956 bps Translation length: 419 residues

Ensembl version ENST00000415254.1

Type Putative protein coding

Prediction Method Manual annotation (determined on a case-by-case basis) from the [Havana](#) project.

Alternative transcripts This transcript corresponds to the following database identifiers:
[Havana transcript: OTTHUMT00000347022](#) (version 1)

Ensembl release 73 - September 2013 © [WTSI](#) / [EBI](#) [About Ensembl](#) | [Privacy Policy](#) | [Contact Us](#)

[Permanent link](#) - [View in archive site](#)

File Edit View History Bookmarks Tools Help

Human genome overview - Genome ... x The Gene Ontology x Ontology Lookup Service (OLS) x Ensembl genome browser 73: Homo ... x +

www.ensembl.org/Homo_sapiens/Transcript/Summary?db=core;g=ENSG00000147576;r=8:66868488-67868487;t=ENST00000415254

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

Search Human...

Human (GRCh37) Location: 8:66,868,488-67,868,487 Gene: ADHFE1 Transcript: ADHFE1-004

Transcript-based displays

- Transcript summary
- Supporting evidence (32)
- Sequence
 - Exons (15)
 - cDNA
 - Protein
- External References
 - General identifiers (21)
 - Oligo probes (29)
- Ontology
 - Ontology graph (10)
 - Ontology table (10)
- Genetic Variation
 - Variation table
 - Variation image
 - Population comparison
 - Comparison image
- Protein Information
 - Protein summary
 - Domains & features (3)
 - Variations (126)
- External data
 - Personal annotation
- ID History
 - Transcript history
 - Protein history

Configure this page

Add your data

Export data

Bookmark this page

Share this page

Transcript: ADHFE1-004 ENST00000415254

Description alcohol dehydrogenase, iron containing, 1 [Source:HGNC Symbol;Acc:16354]

Location [Chromosome 8: 67,344,727-67,381,044](#) forward strand.

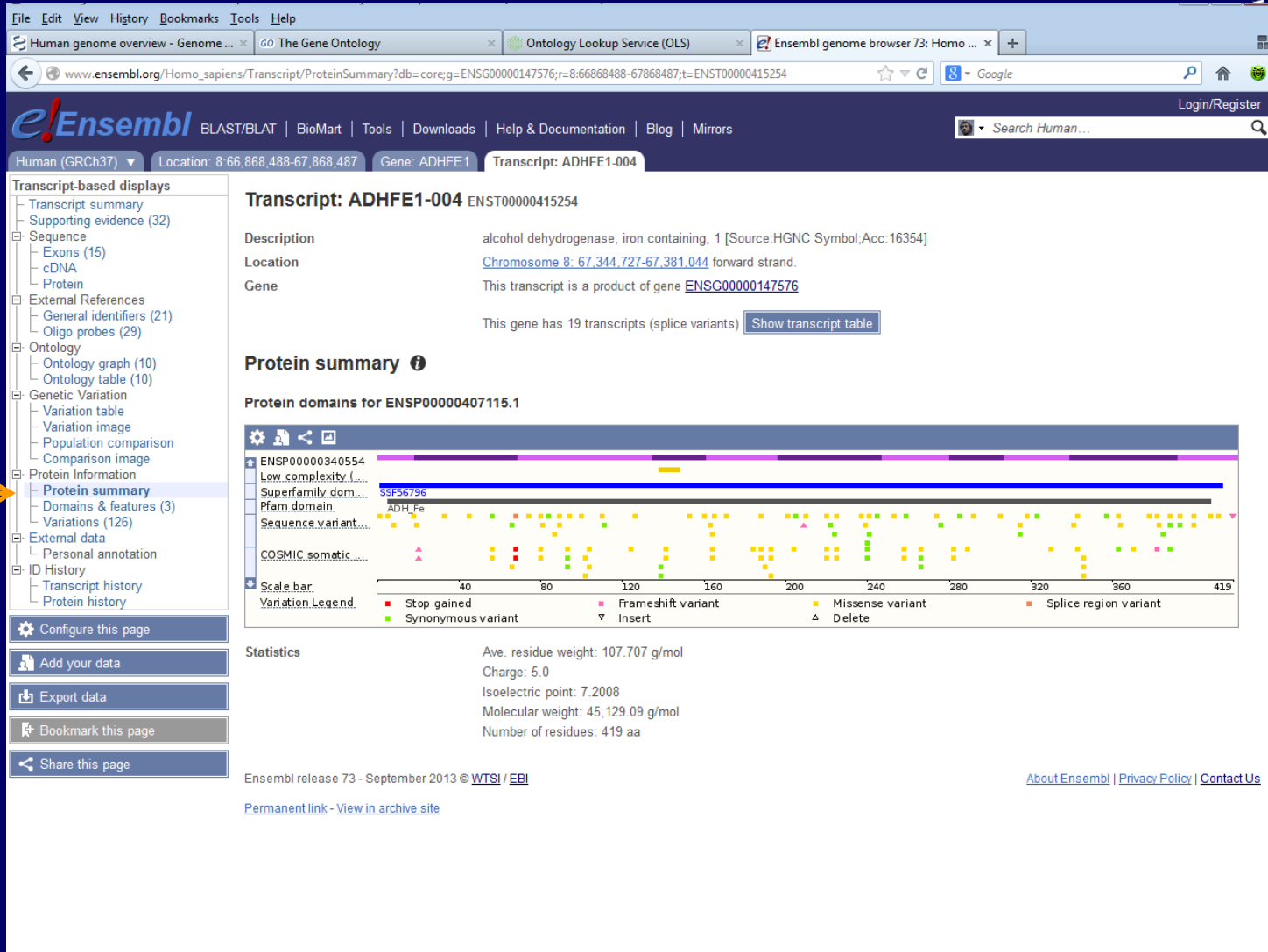
Gene This transcript is a product of gene [ENSG00000147576](#)

This gene has 19 transcripts (splice variants) [Hide transcript table](#)

Show **All** entries Show/hide columns Filter

Name	Transcript ID	Length (bp)	Protein ID	Length (aa)	Biotype	CDS incomplete	CCDS
ADHFE1-001	ENST00000396623	1990	ENSP00000379865	467	Protein coding	-	CCDS6190
ADHFE1-004	ENST00000415254	1956	ENSP00000407115	419	Protein coding	-	-
ADHFE1-201	ENST00000379385	946	ENSP00000368695	303	Protein coding	-	-
ADHFE1-018	ENST00000523113	557	ENSP00000428055	88	Protein coding	3'	-
ADHFE1-002	ENST00000424777	2229	ENSP00000410883	303	Nonsense mediated decay	-	-
ADHFE1-006	ENST00000426810	2179	ENSP00000406905	38	Nonsense mediated decay	-	-
ADHFE1-009	ENST00000419955	1979	ENSP00000392040	34	Nonsense mediated decay	-	-
ADHFE1-003	ENST00000276576	1830	ENSP00000276576	38	Nonsense mediated decay	-	-
ADHFE1-005	ENST00000396621	1006	ENSP00000379864	126	Nonsense mediated decay	-	-
ADHFE1-008	ENST00000422166	982	ENSP00000387956	131	Nonsense mediated decay	-	-
ADHFE1-012	ENST00000449512	556	ENSP00000416450	35	Nonsense mediated decay	-	-
ADHFE1-011	ENST00000431959	524	ENSP00000392925	34	Nonsense mediated decay	-	-
ADHFE1-013	ENST00000443372	515	ENSP00000409714	35	Nonsense mediated decay	-	-
ADHFE1-007	ENST00000496501	1460	No protein product	-	Processed transcript	-	-
ADHFE1-017	ENST00000518781	821	No protein product	-	Processed transcript	-	-
ADHFE1-016	ENST00000466739	609	No protein product	-	Processed transcript	-	-
ADHFE1-015	ENST00000480040	567	No protein product	-	Processed transcript	-	-
ADHFE1-010	ENST00000466920	570	No protein product	-	Retained intron	-	-
ADHFE1-014	ENST00000463261	529	No protein product	-	Retained intron	-	-

Transcript summary ⓘ



The screenshot displays the Ensembl genome browser interface for the transcript ADHFE1-004 (ENST00000415254). The browser window shows the URL www.ensembl.org/Homo_sapiens/Transcript/ProteinSummary?db=core;g=ENSG00000147576;r=8:66868488-67868487;t=ENST00000415254. The left sidebar contains a navigation menu with categories like Transcript-based displays, External References, and Protein Information. The 'Protein summary' link is highlighted with a blue arrow. The main content area shows the transcript details, including its description as 'alcohol dehydrogenase, iron containing, 1', its location on Chromosome 8, and its gene product. Below this, the 'Protein summary' section displays the protein domains for ENSP00000407115.1, including a scale bar and a variation legend. The variation legend identifies various types of variants: Stop gained (red square), Synonymous variant (green square), Frameshift variant (pink square), Missense variant (yellow square), Insert (inverted triangle), and Delete (triangle). The statistics section provides information about the protein, such as its average residue weight, charge, isoelectric point, molecular weight, and number of residues.

Transcript: ADHFE1-004 ENST00000415254

Description alcohol dehydrogenase, iron containing, 1 [Source:HGNC Symbol;Acc:16354]

Location [Chromosome 8: 67,344,727-67,381,044](#) forward strand.

Gene This transcript is a product of gene [ENSG00000147576](#)

This gene has 19 transcripts (splice variants) [Show transcript table](#)

Protein summary ⓘ

Protein domains for ENSP00000407115.1

ENSP00000340554
 Low complexity (.....)
 Superfamily dom.....
 Pfam domain
 Sequence variant.....
 COSMIC somatic.....

Scale bar: 40 80 120 160 200 240 280 320 360 419

Variation Legend:

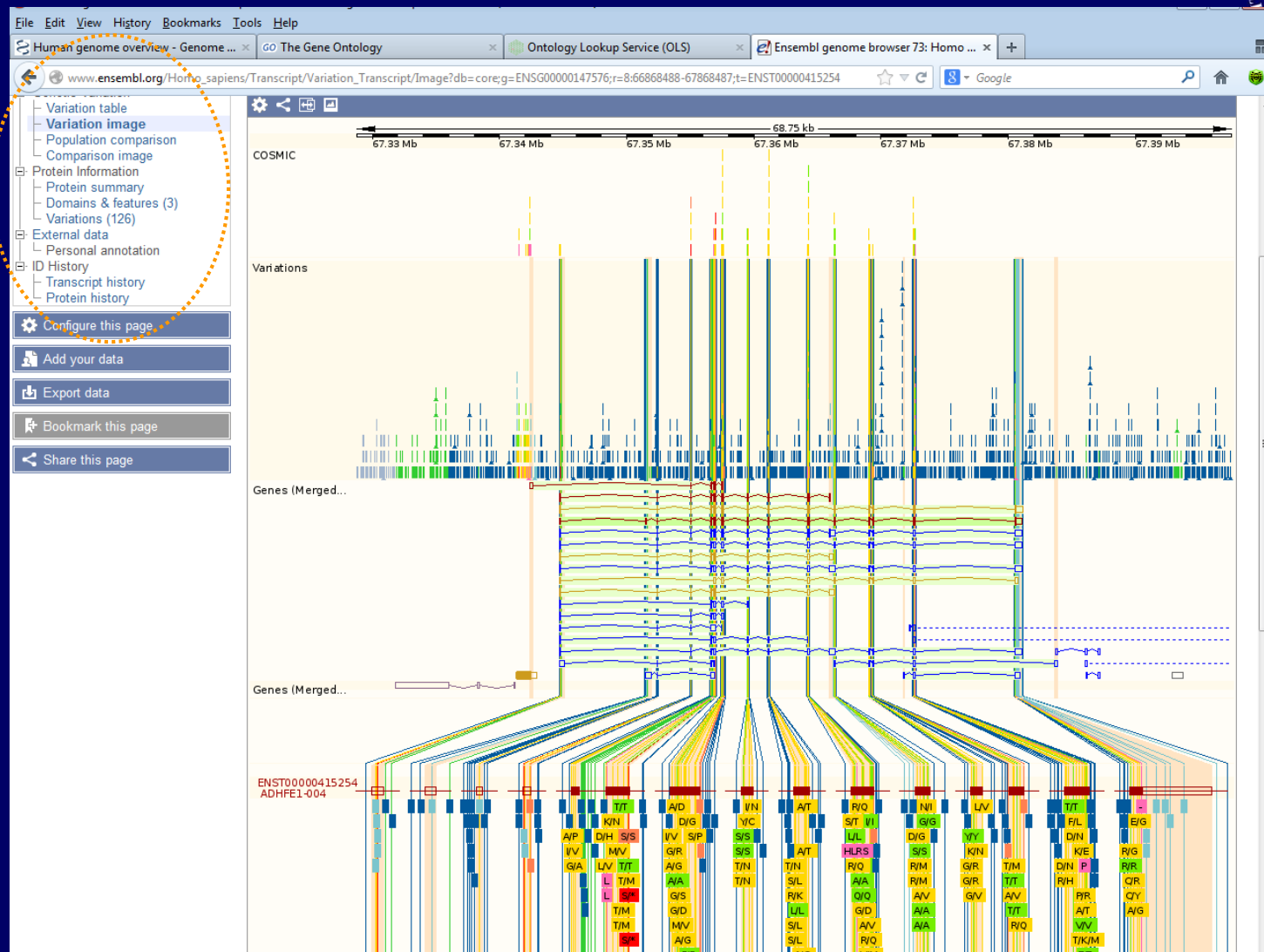
- Stop gained
- Synonymous variant
- Frameshift variant
- Missense variant
- Insert
- Delete
- Splice region variant

Statistics

Ave. residue weight: 107.707 g/mol
 Charge: 5.0
 Isoelectric point: 7.2008
 Molecular weight: 45,129.09 g/mol
 Number of residues: 419 aa

Ensembl release 73 - September 2013 © [WTSI](#) / [EBI](#) [About Ensembl](#) | [Privacy Policy](#) | [Contact Us](#)

[Permanent link - View in archive site](#)



File Edit View History Bookmarks Tools Help

Human genome overview - Genome ... x Go The Gene Ontology x Ontology Lookup Service (OLS) x Ensembl genome browser 73: Homo ... x +

www.ensembl.org/Homo_sapiens/Transcript/Ontology/Table?db=core;g=ENSG00000147576;r=8:66868488-67868487;t=ENST00000415254

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

Human (GRCh37) Location: 8:66,868,488-67,868,487 Gene: ADHFE1 Transcript: ADHFE1-004

Transcript-based displays

- Transcript summary
- Supporting evidence (32)
- Sequence
 - Exons (15)
 - cDNA
 - Protein
- External References
 - General identifiers (21)
 - Oligo probes (29)
- Ontology
 - Ontology graph (10)
 - Ontology table (10)
- Genetic Variation
 - Variation table
 - Variation image
 - Population comparison
 - Comparison image
- Protein Information
 - Protein summary
 - Domains & features (3)
 - Variations (126)
- External data
 - Personal annotation
- ID History
 - Transcript history
 - Protein history

Configure this page

Add your data

Export data

Bookmark this page

Share this page

Download view as CSV

Transcript: ADHFE1-004 ENST00000415254

Description alcohol dehydrogenase, iron containing, 1 [Source:HGNC Symbol;Acc:16354]

Location [Chromosome 8: 67,344,727-67,381,044](#) forward strand.

Gene This transcript is a product of gene [ENSG00000147576](#)

This gene has 19 transcripts (splice variants) [Show transcript table](#)

Ontology table ⓘ

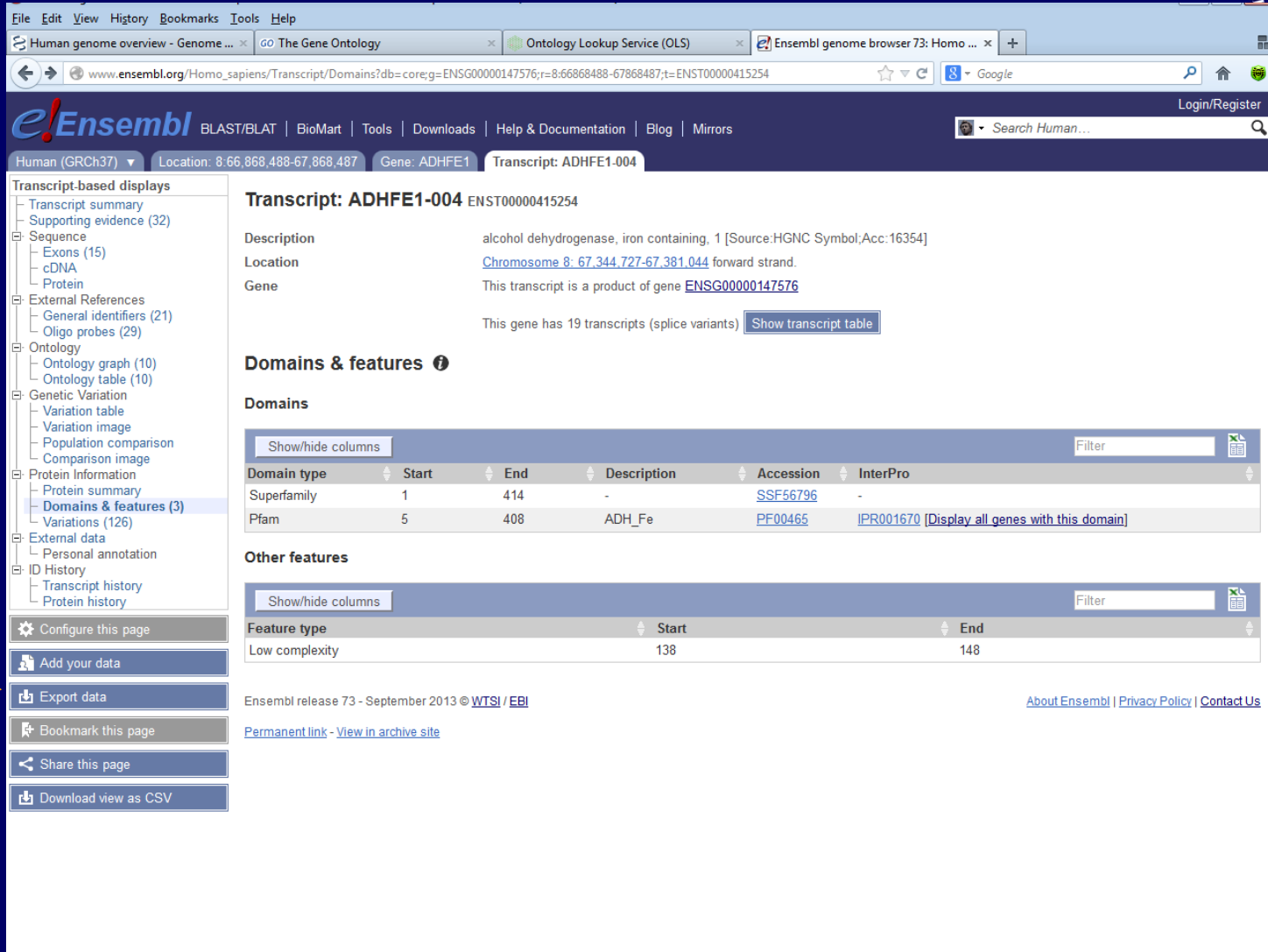
- [GO: Biological process](#)
- [GO: Cellular component](#)
- [GO: Molecular function](#)

Descendants of GO: Biological process

Accession	Term	Evidence	Annotation Source	GOSlim Accessions	GOSlim Terms
GO:0006103	2-oxoglutarate metabolic process	TAS	UniProtKB/Swiss-Prot:HOT_HUMAN	GO:0008150 GO:0044281	biological_process small molecule metabolic process
GO:0015993	molecular hydrogen transport	IDA	UniProtKB/Swiss-Prot:HOT_HUMAN	GO:0006810 GO:0008150	transport biological_process
GO:0044237	cellular metabolic process	TAS	UniProtKB/Swiss-Prot:HOT_HUMAN	GO:0008150	biological_process
GO:0044281	small molecule metabolic process	TAS	UniProtKB/Swiss-Prot:HOT_HUMAN	GO:0008150	biological_process
GO:0055114	oxidation-reduction process	IEA	InterPro:ADH_Fe	GO:0008150	biological_process

Descendants of GO: Cellular component

Accession	Term	Evidence	Annotation Source	GOSlim Accessions	GOSlim Terms
GO:0005739	mitochondrion	IEA	UniProtKB/TrEMBL:E5RFL4_HUMAN	GO:0005575 GO:0043226 GO:0005623 GO:0005622 GO:0005737	cellular_component organelle cell intracellular cytoplasm



Transcript: ADHFE1-004 ENST00000415254

Description alcohol dehydrogenase, iron containing, 1 [Source:HGNC Symbol;Acc:16354]
Location [Chromosome 8: 67,344,727-67,381,044](#) forward strand.
Gene This transcript is a product of gene [ENSG00000147576](#)
 This gene has 19 transcripts (splice variants) [Show transcript table](#)

Domains & features ⓘ

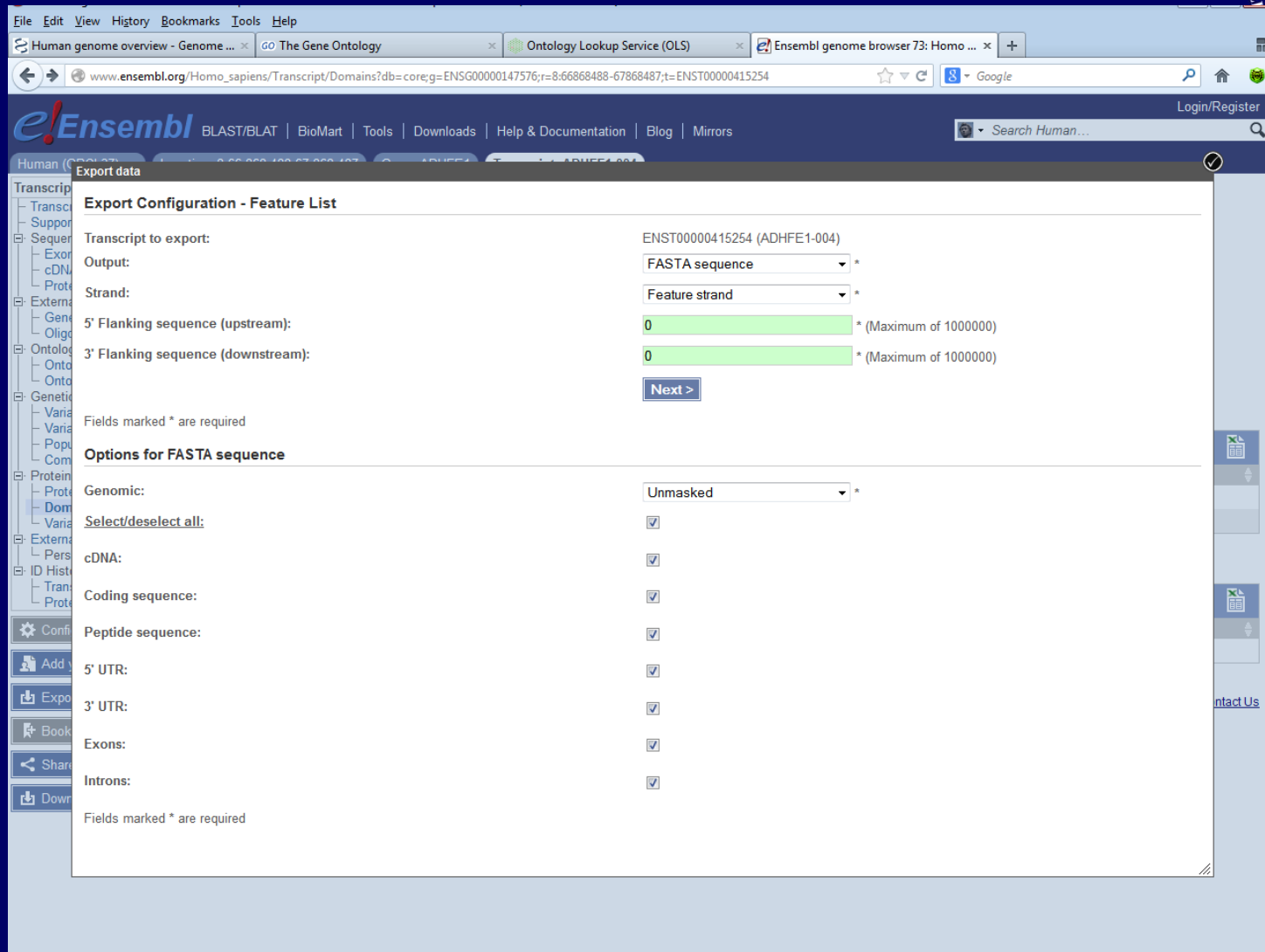
Domains

Domain type	Start	End	Description	Accession	InterPro
Superfamily	1	414	-	SSF56796	-
Pfam	5	408	ADH_Fe	PF00465	IPR001670 [Display all genes with this domain]

Other features

Feature type	Start	End
Low complexity	138	148

Ensembl release 73 - September 2013 © [WTSI](#) / [EBI](#) [About Ensembl](#) | [Privacy Policy](#) | [Contact Us](#)
[Permanent link - View in archive site](#)



File Edit View History Bookmarks Tools Help

Human genome overview - Genome ... The Gene Ontology Ontology Lookup Service (OLS) Ensembl genome browser 73: Homo ...

www.ensembl.org/Homo_sapiens/Transcript/Domains?db=core;g=ENSG00000147576;r=8:66868488-67868487;t=ENST00000415254

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

Search Human ... Login/Register

Human (GRCh37) 8:66,868,488-67,868,487 (ADHFE1)

Export data

Export Configuration - Feature List

Transcript to export: ENST00000415254 (ADHFE1-004)

Output: FASTA sequence *

Strand: Feature strand *

5' Flanking sequence (upstream): 0 * (Maximum of 1000000)

3' Flanking sequence (downstream): 0 * (Maximum of 1000000)

Next >

Fields marked * are required

Options for FASTA sequence

Genomic: Unmasked *

Select/deselect all: ☒

cDNA: ☒

Coding sequence: ☒

Peptide sequence: ☒

5' UTR: ☒

3' UTR: ☒

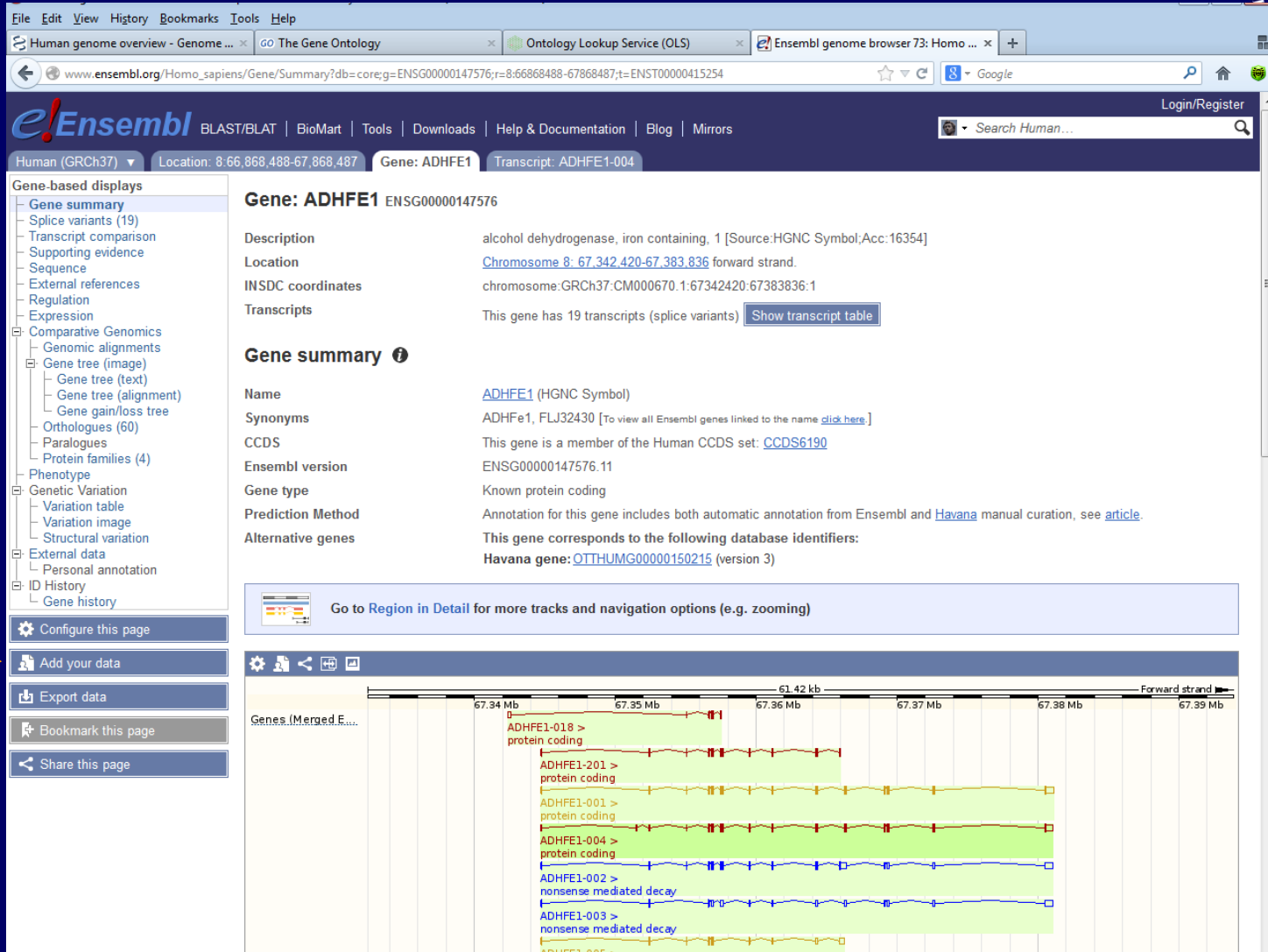
Exons: ☒

Introns: ☒

Fields marked * are required

Az osztott annotációs rendszer (DAS)

- Célja:
 - Saját adatok feltüntetése az ENSEMBL felületen
 - Adatok megosztása együttműködők közt
- Feltöltéshez használható egy külön fül a felületen
- A formátum kötött, alapvetően szöveges (ld. „Help”)
- ENSEMBL regisztráció és belépés segít



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

Human (GRCh37) Location: 8:66,868,488-67,868,487 Gene: **ADHFE1** Transcript: ADHFE1-004

Gene-based displays

- Gene summary
- Splice variants (19)
- Transcript comparison
- Supporting evidence
- Sequence
- External references
- Regulation
- Expression
- Comparative Genomics
 - Genomic alignments
 - Gene tree (image)
 - Gene tree (text)
 - Gene tree (alignment)
 - Gene gain/loss tree
 - Orthologues (60)
 - Paralogues
 - Protein families (4)
- Phenotype
- Genetic Variation
 - Variation table
 - Variation image
 - Structural variation
- External data
 - Personal annotation
- ID History
- Gene history

Gene: ADHFE1 ENSG00000147576

Description alcohol dehydrogenase, iron containing, 1 [Source:HGNC Symbol;Acc:16354]

Location [Chromosome 8: 67,342,420-67,383,836](#) forward strand.

INSDC coordinates chromosome:GRCh37:CM000670.1:67342420-67383836:1

Transcripts This gene has 19 transcripts (splice variants) [Show transcript table](#)

Gene summary

Name [ADHFE1](#) (HGNC Symbol)

Synonyms ADHFE1, FLJ32430 [To view all Ensembl genes linked to the name [click here](#)]

CCDS This gene is a member of the Human CCDS set: [CCDS6190](#)

Ensembl version ENSG00000147576.11

Gene type Known protein coding

Prediction Method Annotation for this gene includes both automatic annotation from Ensembl and [Havana](#) manual curation, see [article](#).

Alternative genes This gene corresponds to the following database identifiers:
Havana gene: [QTTHUMG00000150215](#) (version 3)

[Go to Region in Detail](#) for more tracks and navigation options (e.g. zooming)

Genes (Merged E...)

67.34 Mb 67.35 Mb 67.36 Mb 67.37 Mb 67.38 Mb 67.39 Mb

61.42 kb Forward strand

ADHFE1-018 > protein coding

ADHFE1-201 > protein coding

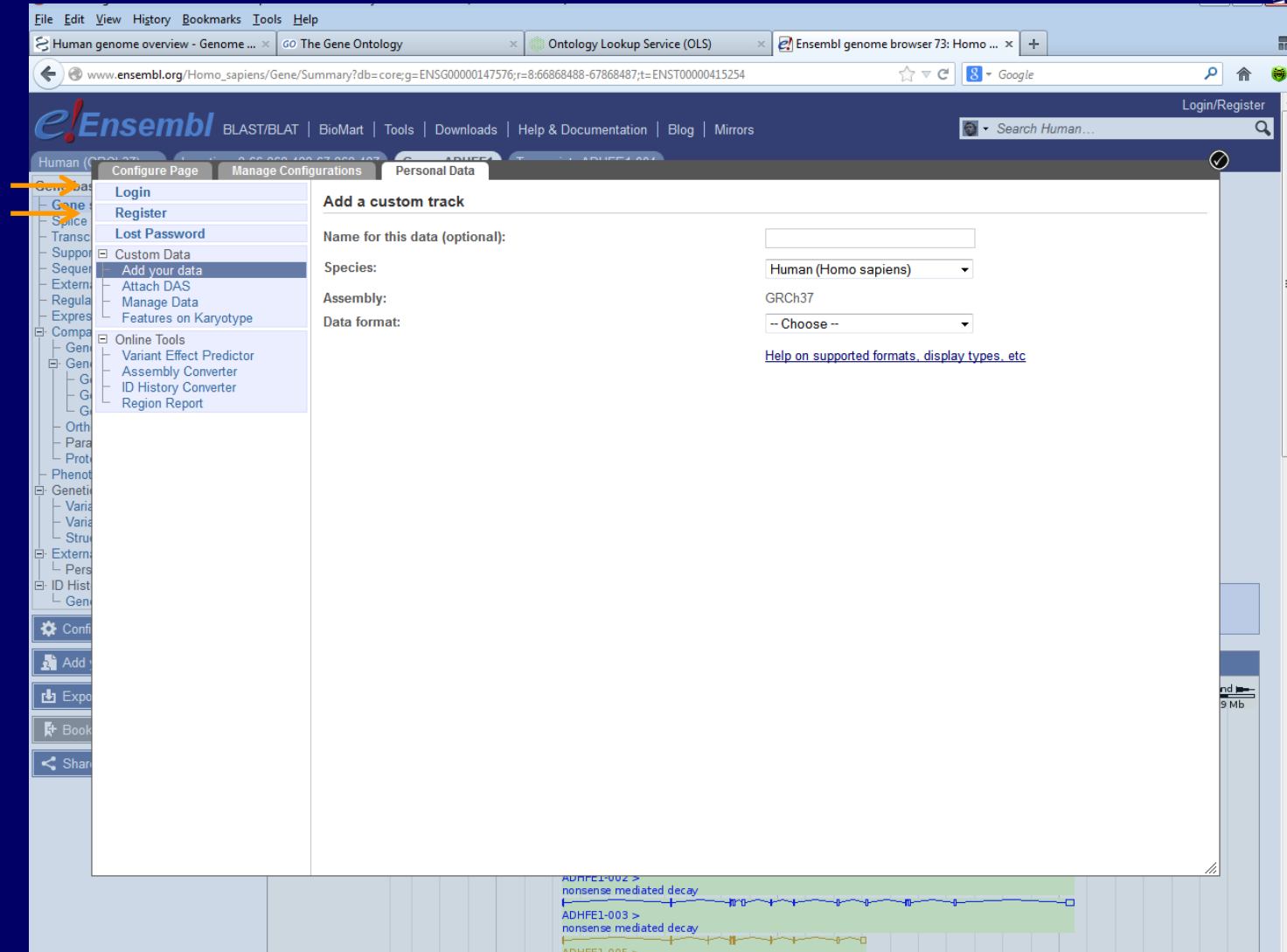
ADHFE1-001 > protein coding

ADHFE1-004 > protein coding

ADHFE1-002 > nonsense mediated decay

ADHFE1-003 > nonsense mediated decay

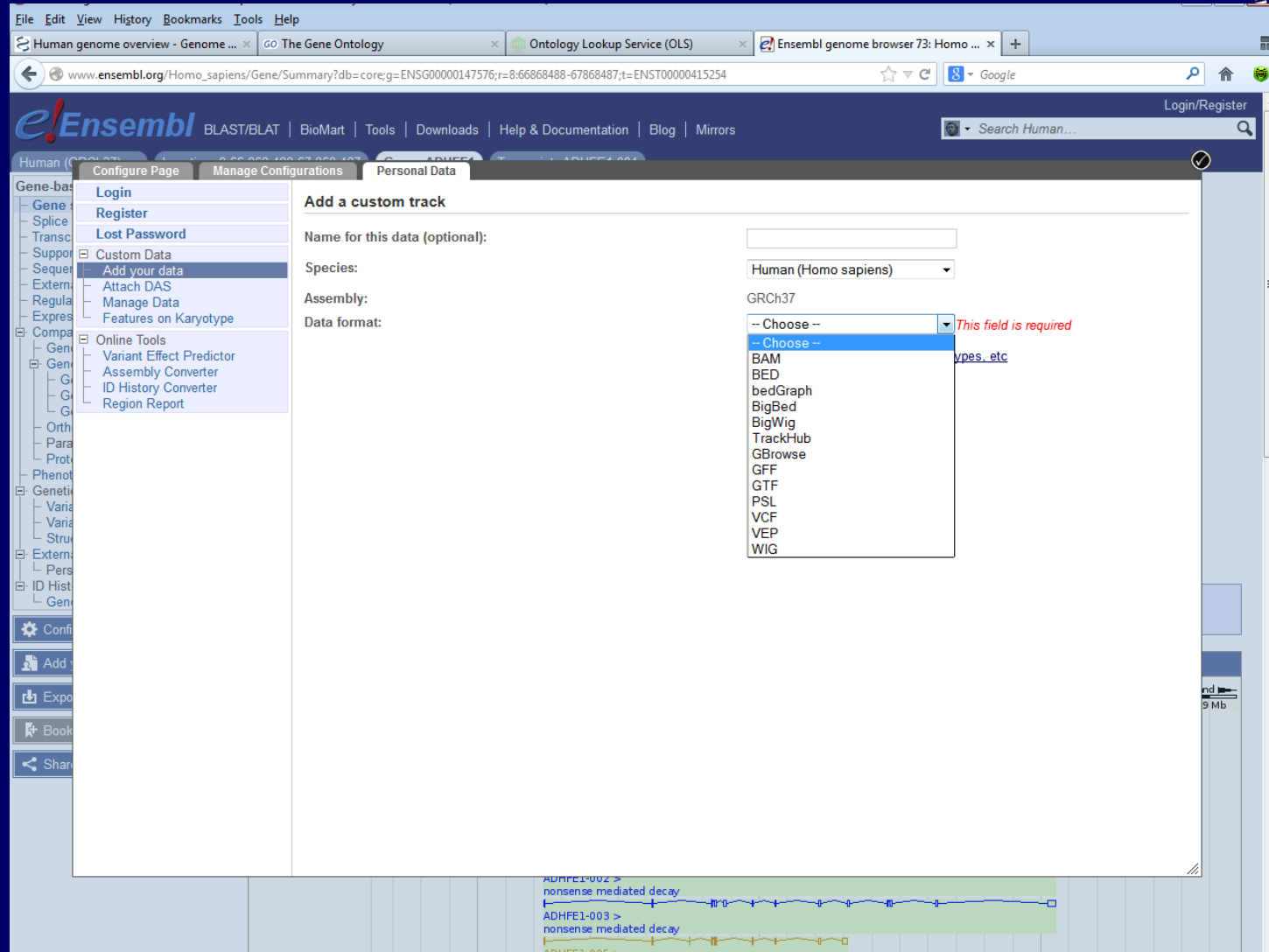
ADHFE1-005 >



The screenshot displays the Ensembl genome browser interface. The browser window shows the URL `www.ensembl.org/Homo_sapiens/Gene/Summary?db=core;g=ENSG00000147576;r=8:66868488-67868487;t=ENST00000415254`. The left sidebar contains a navigation menu with categories like 'Gene', 'Transcript', 'Support', 'Sequences', 'External', 'Regulation', 'Expression', 'Comparative', 'Genetics', 'Phenotype', 'Genetic', 'Variation', 'Structure', 'External', 'Personal', 'ID History', and 'Gene'. The 'Custom Data' option is selected, and the 'Add your data' sub-option is highlighted. The main content area is titled 'Add a custom track' and contains the following fields:

- Name for this data (optional):
- Species:
- Assembly:
- Data format:

A link [Help on supported formats, display types, etc](#) is provided below the data format field. The bottom of the browser shows a genomic track with labels like 'ADHFE1-002 > nonsense mediated decay' and 'ADHFE1-003 > nonsense mediated decay'.



The screenshot displays the Ensembl genome browser interface. The main content area shows the 'Add a custom track' dialog box. The 'Species' is set to 'Human (Homo sapiens)' and the 'Assembly' is 'GRCh37'. The 'Data format' dropdown menu is open, showing a list of file formats: BAM, BED, bedGraph, BigBed, BigWig, TrackHub, GBrowse, GFF, GTF, PSL, VCF, VEP, and WIG. A red error message 'This field is required' is displayed next to the 'Data format' dropdown. The background shows the Ensembl website with a sidebar menu and a track visualization at the bottom.

Kapcsolódó eszközök

- BLAST/BLAT
- BioMart
- Adatletöltési lehetőségek

File Edit View History Bookmarks Tools Help

Human genome overview - Genome ... x The Gene Ontology x Ontology Lookup Service (OLS) x BLAST Search x +

www.ensembl.org/Homo_sapiens/blastview

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

Search Human...

new **SETUP** ↔ CONFIG ↔ RESULTS ↔ DISPLAY **refresh** **Online Help**

Important Notice
We now use Blat as our default DNA search. This will make your query faster.

Enter the Query Sequence
Either Paste sequences (max 30 sequences) in FASTA or plain text:

Or Upload a file containing one or more FASTA sequences
 No file selected.

Or Enter a sequence ID or accession (EMBL, UniProt, RefSeq)

Or Enter an existing ticket ID:

☒ dna queries
☐ peptide queries

Select the databases to search against
Select species:
Use 'ctrl' key to select multiple species
Gasterosteus_aculeatus
Gorilla_gorilla
Homo_sapiens

☒ dna database
☐ protein database

Genomic sequence
Proteins

Select the Search Tool
BLASTN
BLAT
TBLASTX

configure ▶ **RUN**

Search sensitivity:
Optimise search parameters to find the following alignments
Near-exact matches

Summary
setup
▶ Not yet initialised
configure
▶ Not yet initialised
results
▶ Not yet initialised
display
▶ Not yet initialised

File Edit View History Bookmarks Tools Help

Human genome overview - Genome... x The Gene Ontology x Ontology Lookup Service (OLS) x BLAST Search x Ensembl genome browser 73: Homo... x +

www.ensembl.org/Homo_sapiens/blastview/BLA_KM6c5pq2L

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

Search Human...

new **SETUP** **CONFIG** **RESULTS** **DISPLAY**

refresh Online Help

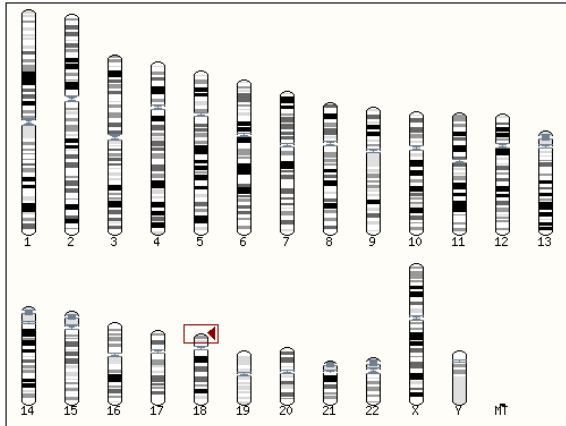
Summary

- setup
 - Homo_sapiens
 - Genomic sequence
 - BLAT
 - Exact sensitivity
- configure
- results
- display
 - Not yet initialised

Displaying Icl18_514562_515316 sequence alignments vs Homo_sapiens LATESTGP database

Showing top 100 alignments of 22, sorted by Raw Score

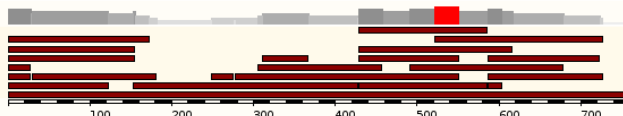
☒ Alignment Locations vs. Karyotype (click arrow to hide)



☒ Alignment Locations vs. Query (click arrow to hide)

coverage

HSPs



☒ Alignment Summary (click arrow to hide)

Select rows to include in table, and type of sort
(Use the 'ctrl' key to select multiples)

refresh

File Edit View History Bookmarks Tools Help

Human genome overview - Genome... x The Gene Ontology x Ontology Lookup Service (OLS) x BLAST Search x Ensembl genome browser 73: Homo... x

www.ensembl.org/Homo_sapiens/blastview/BLA_KM6c5pq2L

☒ Alignment Locations vs. Query (click arrow to hide)

coverage

☒ Alignment Summary (click arrow to hide)

Select rows to include in table, and type of sort
(Use the 'ctrl' key to select multiples)

[refresh](#)

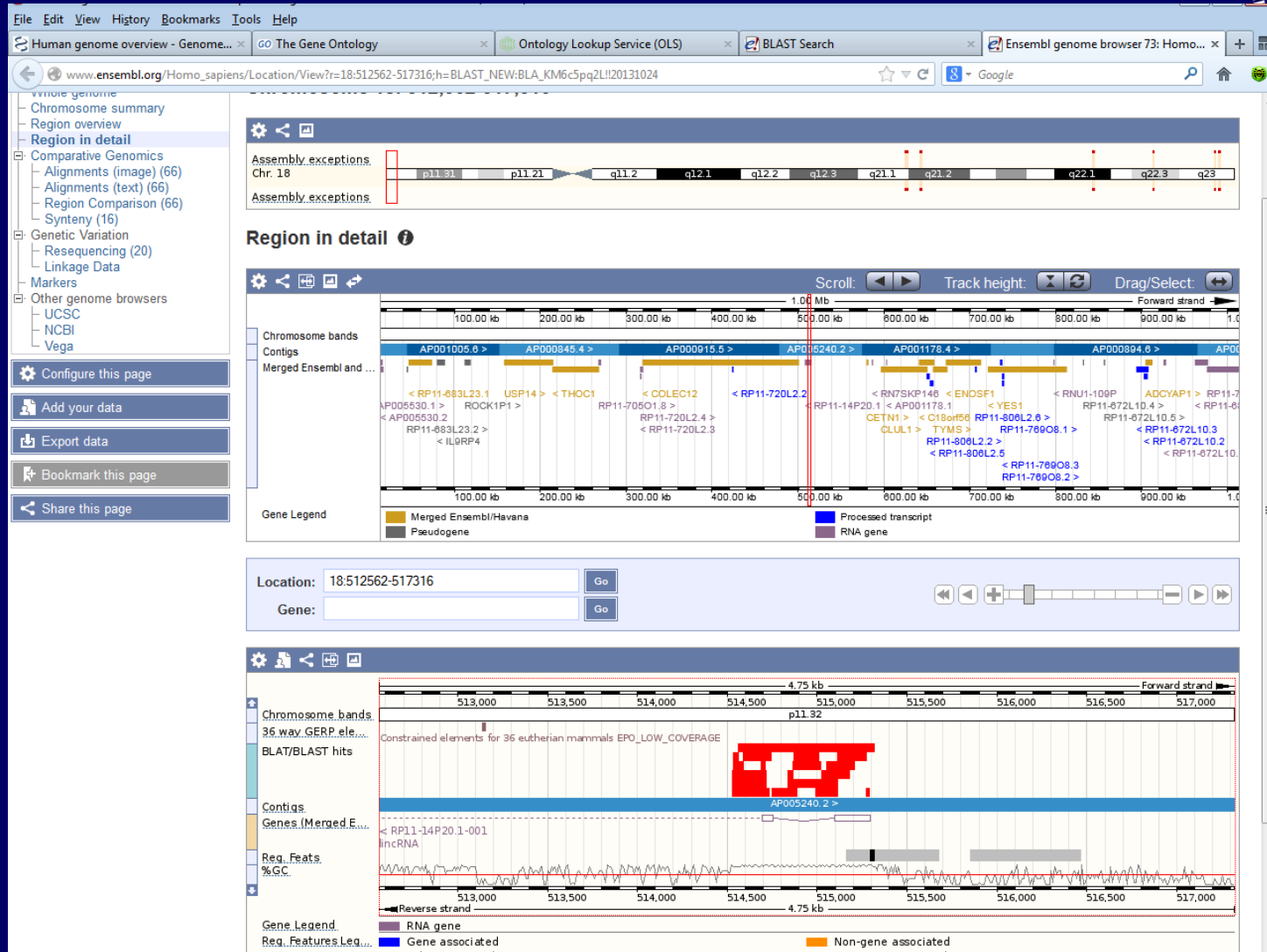
Query	Subject	Chromosome	Supercontig	Clone	Contig	Lrg	Stats	Sort By
off	_off_	_off_	_off_	_off_	_off_	_off_	_off_	>Lrg
Name	Name	Name	Name	Name	Name	Name	Score	<Score
Start	Start	Start	Start	Start	Start	Start	E-val	>Score
Links	Query	Chromosome	Supercontig	Clone	Contig	Lrg	Stats	
	Start End Ori	Name Start End Ori					Score E-val %ID Length	
[A] [G] [C]	1 755 +	Chr:18 514562 515316 +					3886 0.0e+00 100.00 755	
[A] [G] [C]	153 427 +	Chr:18 514837 515111 +					1414 1.4e-242 100.00 275	
[A] [G] [C]	278 550 +	Chr:18 514716 514988 +					1404 1.1e-240 100.00 273	
[A] [G] [C]	522 726 +	Chr:18 514529 514733 +					1021 9.1e-173 97.56 205	
[A] [G] [C]	429 616 +	Chr:18 514529 514716 +					953 1.0e-160 98.94 188	
[A] [G] [C]	491 678 +	Chr:18 514529 514716 +					941 2.0e-158 97.87 188	
[A] [G] [C]	1 172 +	Chr:18 515116 515287 +					858 8.3e-144 97.67 172	
[A] [G] [C]	429 585 +	Chr:18 514560 514716 +					807 9.1e-135 100.00 157	
[A] [G] [C]	1 155 +	Chr:18 515023 515177 +					790 9.3e-132 99.35 155	
[A] [G] [C]	1 155 +	Chr:18 515054 515208 +					784 1.3e-130 98.71 155	
[A] [G] [C]	429 585 +	Chr:18 515114 515270 +					782 3.3e-130 97.45 157	
[A] [G] [C]	306 457 +	Chr:18 514591 514742 +					782 3.6e-130 100.00 152	
[A] [G] [C]	30 181 +	Chr:18 514867 515018 +					782 3.6e-130 100.00 152	
[A] [G] [C]	587 726 +	Chr:18 514747 514886 +					700 1.1e-115 97.86 140	
[A] [G] [C]	587 723 +	Chr:18 514870 515006 +					685 5.6e-113 97.81 137	
[A] [G] [C]	429 550 +	Chr:18 514591 514712 +					627 1.0e-102 100.00 122	
[A] [G] [C]	1 122 +	Chr:18 514992 515113 +					626 1.2e-102 100.00 122	
[A] [G] [C]	311 367 +	Chr:18 514534 514590 +					286 4.7e-42 98.25 57	
[A] [G] [C]	1 27 +	Chr:18 514686 514712 +					138 6.6e-16 100.00 27	
[A] [G] [C]	1 27 +	Chr:18 514839 514865 +					138 6.6e-16 100.00 27	
[A] [G] [C]	249 275 +	Chr:18 514533 514559 +					131 1.1e-14 96.30 27	
[A] [G] [C]	587 603 +	Chr:18 515271 515287 +					87 1.0e-06 100.00 17	

Ensembl release 73 - September 2013 © WTSI / EBI

[About Ensembl](#) | [Privacy Policy](#) | [Contact Us](#)

[Permanent link - View in archive site](#)

www.ensembl.org



Tömeges lekérdezés

- BioMart rendszer
- Web-felületen át elérhető
- Az adattartalom azonos az interaktív felületem megjelenítéssel
- Szöveges eredményt ad
- Nagy adattömeget fogunk kapni eredményül

Human genome overview ... x AmiGO 2: Welcome x Ontology Lookup Service (... x http://www.e...97c9816967f4 x +

www.ensembl.org/biomart/martview/39fe1f97d1569ca10c6c97c9816967f4

Google

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

Login/Register

Search all species...

New Count Results

URL XML Perl Help

Dataset

[None selected]

- CHOOSE DATABASE -

- CHOOSE DATABASE -

Ensembl Genes 77

Ensembl Variation 77

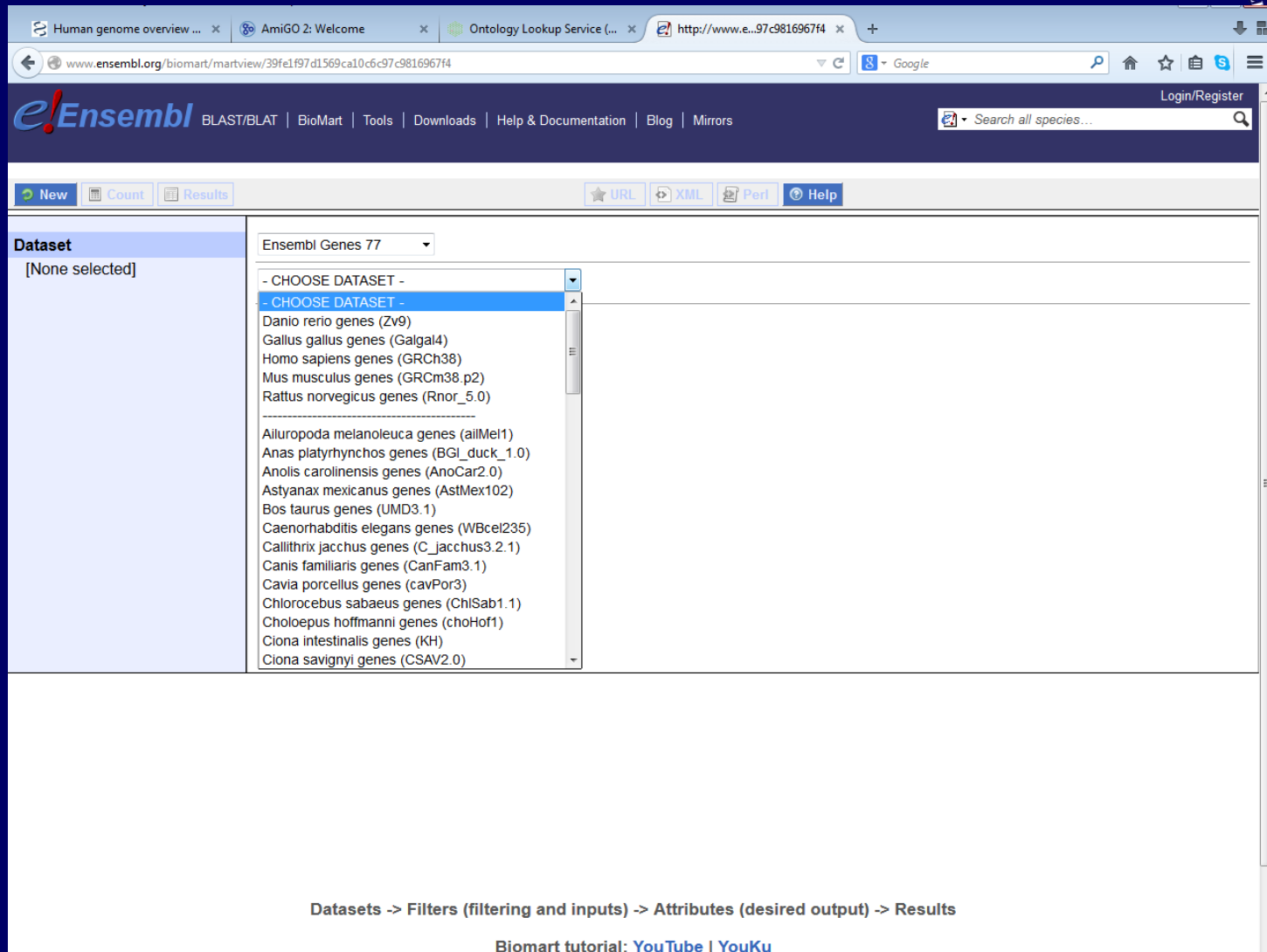
Ensembl Regulation 77

Vega 57

PRIDE (EBI UK)

Datasets -> Filters (filtering and inputs) -> Attributes (desired output) -> Results

Biomart tutorial: [YouTube](#) | [YouKu](#)



Human genome overview ... x AmiGO 2: Welcome x Ontology Lookup Service (... x http://www.e...97c9816967f4 x +

www.ensembl.org/biomart/martview/39fe1f97d1569ca10c6c97c9816967f4

Google

Login/Register

Search all species...

New Count Results URL XML Perl Help

Dataset

[None selected]

Ensembl Genes 77

- CHOOSE DATASET -

- CHOOSE DATASET -

Danio rerio genes (Zv9)

Gallus gallus genes (GallGal4)

Homo sapiens genes (GRCh38)

Mus musculus genes (GRCm38.p2)

Rattus norvegicus genes (Rnor_5.0)

Alluopoda melanoleuca genes (allMel1)

Anas platyrhynchos genes (BGI_duck_1.0)

Anolis carolinensis genes (AnoCar2.0)

Astyanax mexicanus genes (AstMex102)

Bos taurus genes (UMD3.1)

Caenorhabditis elegans genes (WBcel235)

Callithrix jacchus genes (C_jacchus3.2.1)

Canis familiaris genes (CanFam3.1)

Cavia porcellus genes (cavPor3)

Chlorocebus sabaeus genes (ChlSab1.1)

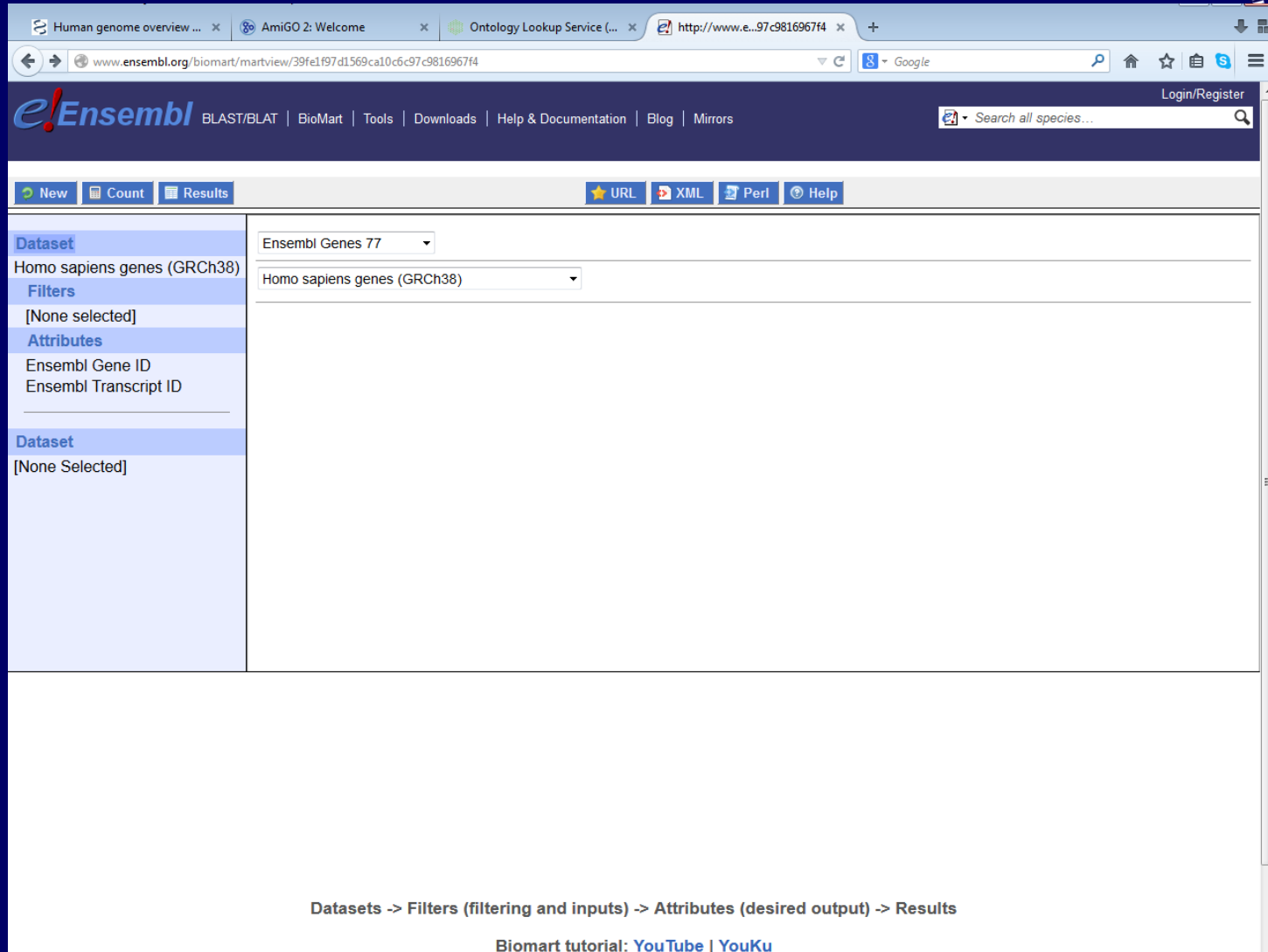
Choloepus hoffmanni genes (choHof1)

Ciona intestinalis genes (KH)

Ciona savignyi genes (CSAV2.0)

Datasets -> Filters (filtering and inputs) -> Attributes (desired output) -> Results

Biomart tutorial: [YouTube](#) | [YouKu](#)



Human genome overview ... x AmiGO 2: Welcome x Ontology Lookup Service (... x http://www.e...97c9816967f4 x

www.ensembl.org/biomart/martview/39fe1f97d1569ca10c6c97c9816967f4

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

Login/Register

Search all species...

New Count Results URL XML Perl Help

Dataset
Homo sapiens genes (GRCh38)

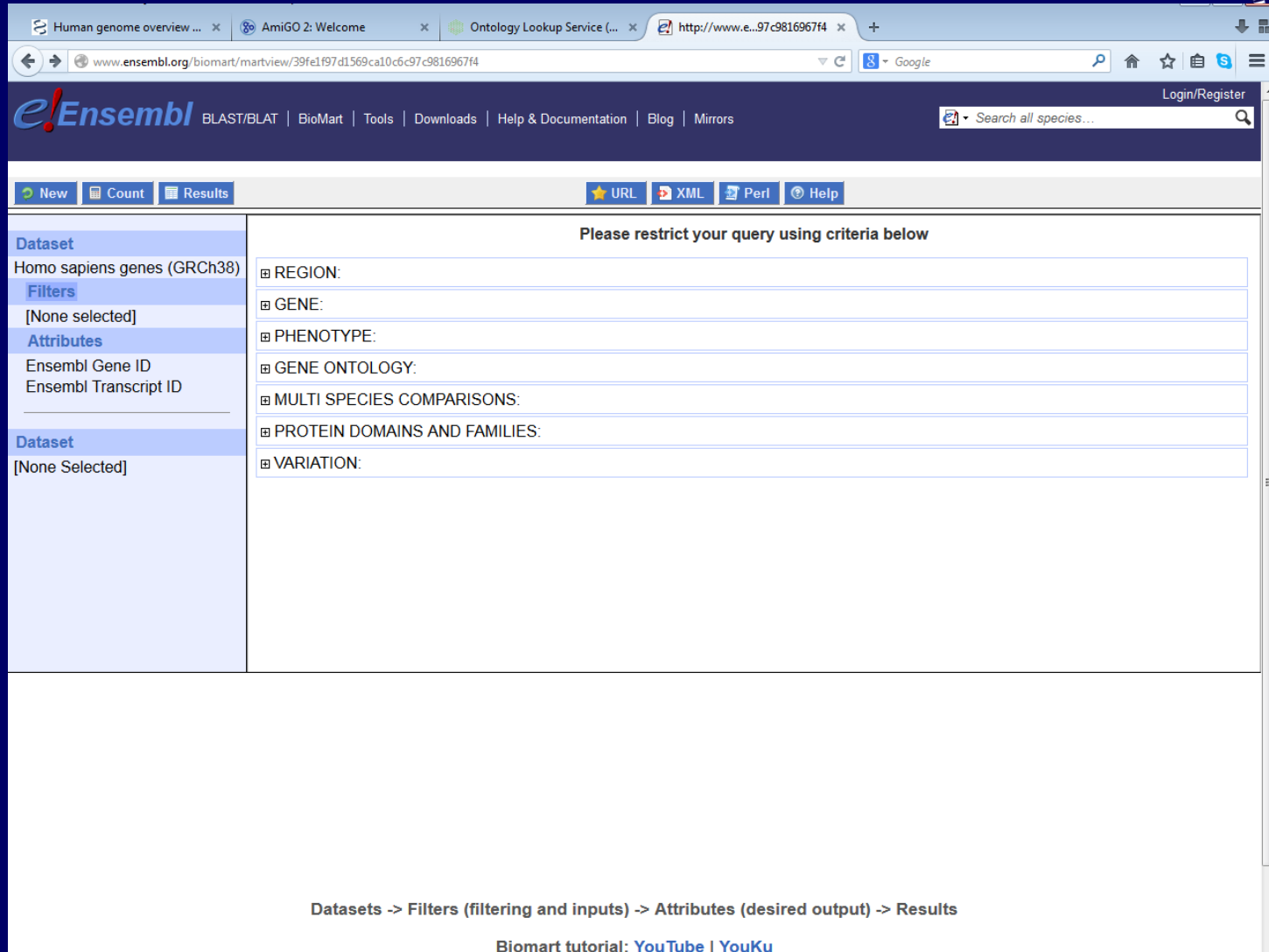
Filters
[None selected]

Attributes
Ensembl Gene ID
Ensembl Transcript ID

Dataset
[None Selected]

Datasets -> Filters (filtering and inputs) -> Attributes (desired output) -> Results

Biomart tutorial: [YouTube](#) | [YouKu](#)



The screenshot shows the Ensembl BioMart interface. The browser tabs include 'Human genome overview ...', 'AmiGO 2: Welcome', 'Ontology Lookup Service (...)', and 'http://www.e...97c9816967f4'. The address bar shows 'www.ensembl.org/biomart/martview/39fe1f97d1569ca10c6c97c9816967f4'. The Ensembl logo is at the top left, with navigation links: BLAST/BLAT, BioMart, Tools, Downloads, Help & Documentation, Blog, Mirrors. A search bar says 'Search all species...'. Below the navigation bar are buttons: New, Count, Results, URL, XML, Perl, Help. The main content area is titled 'Please restrict your query using criteria below'. On the left, there are sections for 'Dataset' (Homo sapiens genes (GRCh38)), 'Filters' ([None selected]), 'Attributes' (Ensembl Gene ID, Ensembl Transcript ID), and another 'Dataset' section ([None Selected]). The main area contains several filter options, each with a checkbox and a text input field: REGION, GENE, PHENOTYPE, GENE ONTOLOGY, MULTI SPECIES COMPARISONS, PROTEIN DOMAINS AND FAMILIES, and VARIATION. At the bottom, a workflow diagram shows 'Datasets -> Filters (filtering and inputs) -> Attributes (desired output) -> Results'. Below this is a link: 'Biomart tutorial: YouTube | YouKu'.

Human genome overview ... x AmiGO 2: Welcome x Ontology Lookup Service (... x http://www.e...97c9816967f4 x +

www.ensembl.org/biomart/martview/39fe1f97d1569ca10c6c97c9816967f4

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

Login/Register

Search all species...

New Count Results

URL XML Perl Help

Dataset
Homo sapiens genes (GRCh38)

Filters
[None selected]
Attributes
Ensembl Gene ID
Ensembl Transcript ID

Dataset
[None Selected]

Please restrict your query using criteria below

☐ REGION:

☐ GENE:

☐ Limit to genes ... with ArrayExpress ID(s) ☒ Only ☐ Excluded

☐ ID list limit [Max 500 advised] Ensembl Gene ID(s) [e.g. ENSG00000139618]

Browse... No file selected.

☐ Transcript count >=

☐ Gene type

3prime_overlapping_ncrna
antisense
IG_C_gene
IG_C_pseudogene
IG_D_gene
IG_J_gene
IG_J_pseudogene
IG_V_gene

☐ Source (gene) ensembl

☐ Source (transcript) ensembl

☐ Status (gene) KNOWN

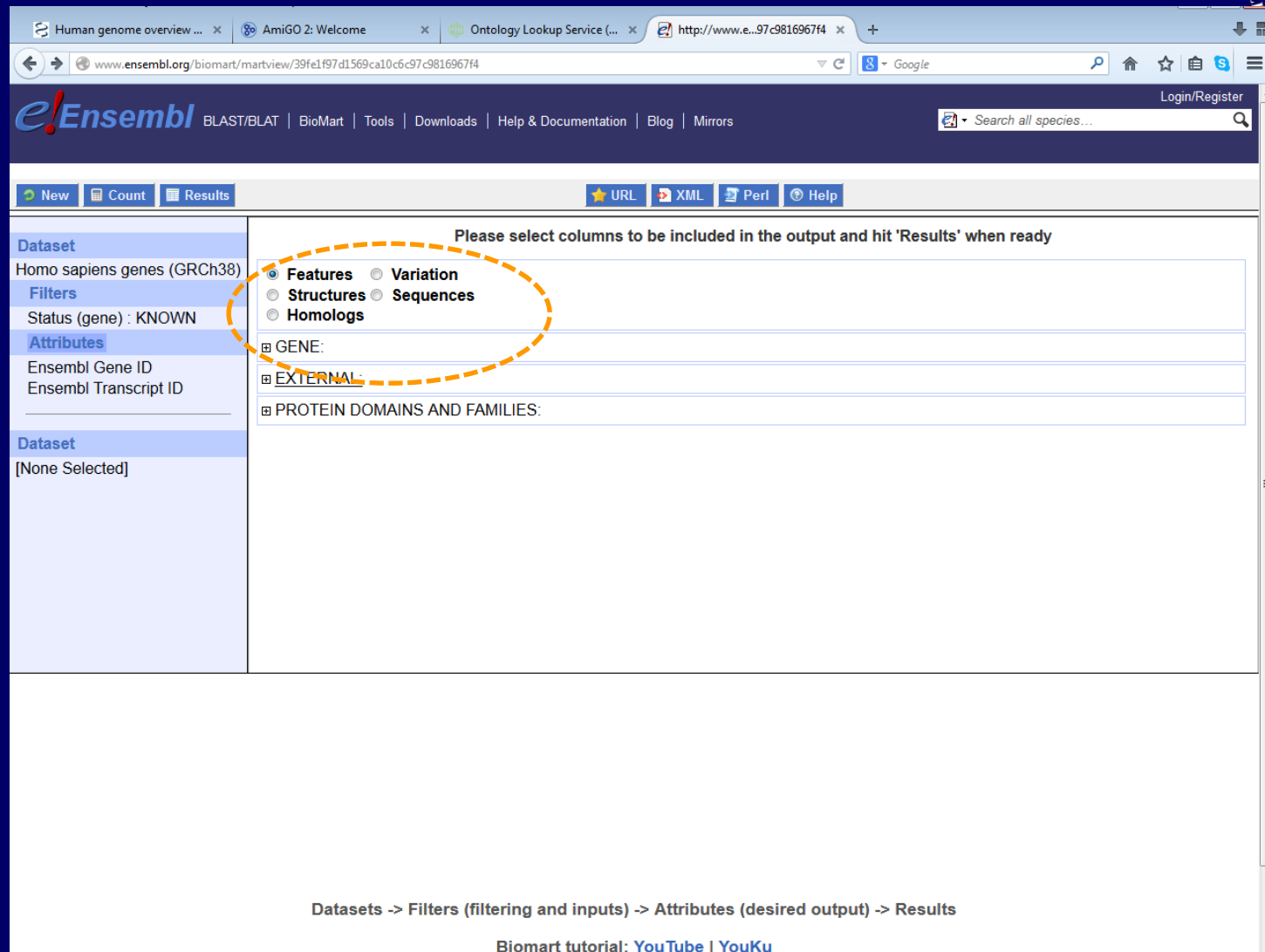
☐ Status (transcript) KNOWN

☐ PHENOTYPE:

☐ GENE ONTOLOGY:

☐ MULTI SPECIES COMPARISONS:

☐ PROTEIN DOMAINS AND FAMILIES:



Human genome overview ... x AmiGO 2: Welcome x Ontology Lookup Service (... x http://www.e...97c9816967f4 x +

www.ensembl.org/biomart/martview/39fe1f97d1569ca10c6c97c9816967f4

Google

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

Login/Register

Search all species...

New Count Results URL XML Perl Help

Dataset
Homo sapiens genes (GRCh38)

Filters
Status (gene) : KNOWN

Attributes
Ensembl Gene ID
Ensembl Transcript ID

Dataset
[None Selected]

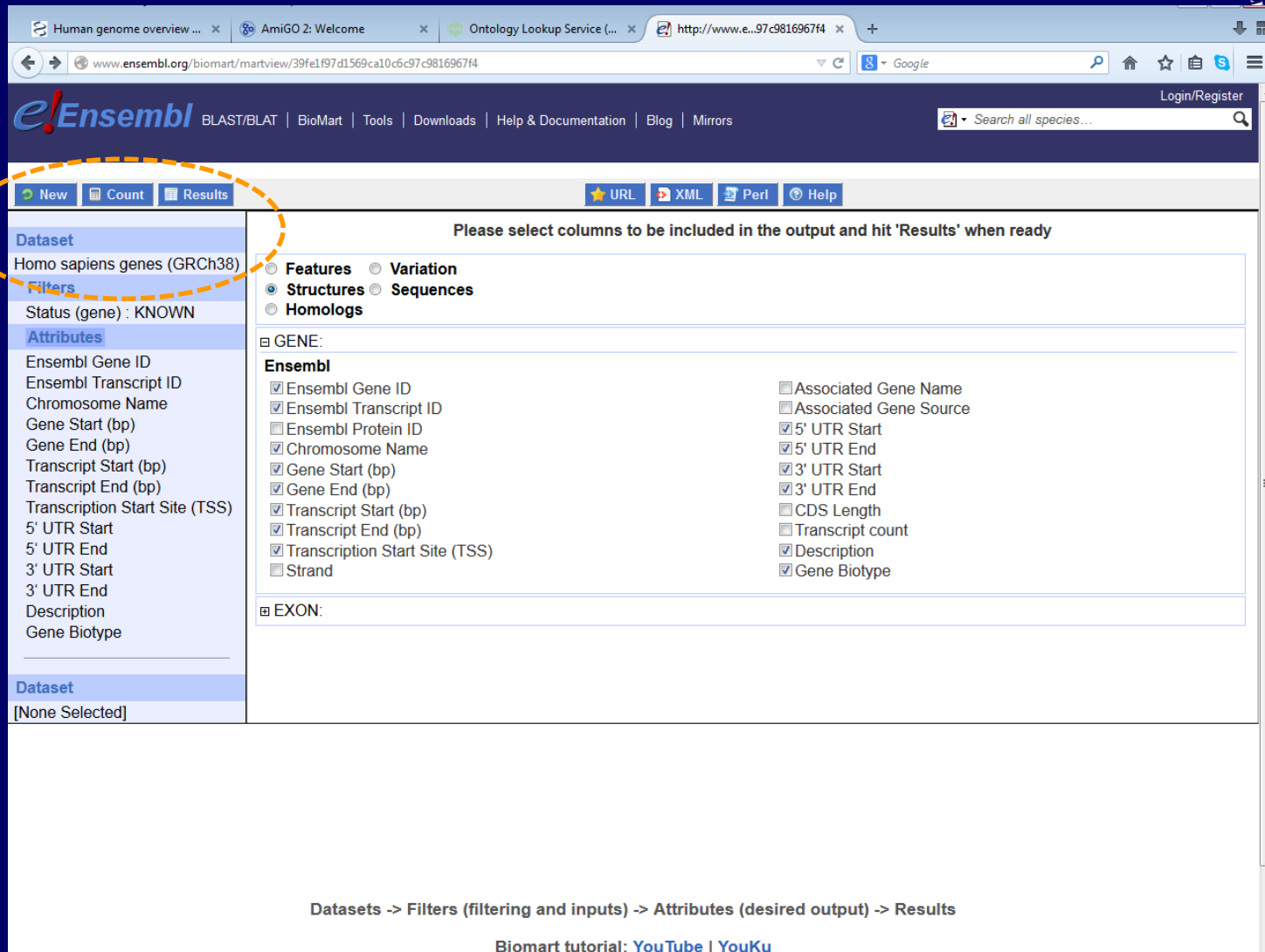
Please select columns to be included in the output and hit 'Results' when ready

☒ Features ☐ Variation
☐ Structures ☐ Sequences
☐ Homologs

☐ GENE:
☐ EXTERNAL:
☐ PROTEIN DOMAINS AND FAMILIES:

Datasets -> Filters (filtering and inputs) -> Attributes (desired output) -> Results

Biomart tutorial: [YouTube](#) | [YouKu](#)



Human genome overview ... x AmiGO 2: Welcome x Ontology Lookup Service (... x http://www.e...97c9816967f4 x +

www.ensembl.org/biomart/martview/39fe1f97d1569ca10c6c97c9816967f4

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

Login/Register

Search all species...

New Count Results

URL XML Perl Help

Please select columns to be included in the output and hit 'Results' when ready

Dataset

Homo sapiens genes (GRCh38)

Filters

Status (gene) : KNOWN

Attributes

Ensembl Gene ID

Ensembl Transcript ID

Chromosome Name

Gene Start (bp)

Gene End (bp)

Transcript Start (bp)

Transcript End (bp)

Transcription Start Site (TSS)

5' UTR Start

5' UTR End

3' UTR Start

3' UTR End

Description

Gene Biotype

Dataset

[None Selected]

Features **Variation**

Structures **Sequences**

Homologs

GENE:

Ensembl

☒ Ensembl Gene ID

☒ Ensembl Transcript ID

☐ Ensembl Protein ID

☒ Chromosome Name

☒ Gene Start (bp)

☒ Gene End (bp)

☒ Transcript Start (bp)

☒ Transcript End (bp)

☒ Transcription Start Site (TSS)

☐ Strand

☐ Associated Gene Name

☐ Associated Gene Source

☒ 5' UTR Start

☒ 5' UTR End

☒ 3' UTR Start

☒ 3' UTR End

☐ CDS Length

☐ Transcript count

☒ Description

☒ Gene Biotype

EXON:

Datasets -> Filters (filtering and inputs) -> Attributes (desired output) -> Results

Biomart tutorial: [YouTube](#) | [YouKu](#)

Human genome overview ... x AmiGO 2: Welcome x Ontology Lookup Service (... x http://www.e...97c9816967f4 x +

www.ensembl.org/biomart/martview/39fe1f97d1569ca10c6c97c9816967f4

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

Login/Register

Search all species...

New Count Results

URL XML Perl Help

Export all results to File TSV Unique results only Go

Email notification to

View 10 rows as HTML Unique results only

Ensembl Gene ID	Ensembl Transcript ID	Chromosome Name	Gene Start (bp)	Gene End (bp)	Transcript Start (bp)	Transcript End (bp)	Transcription Start Site (TSS)	5' UTR Start
ENSG00000276385	ENST00000618935	20	8934984	8935481	8934984	8935481	8935481	
ENSG00000197468	ENST00000508957	4	8949630	8950559	8949630	8950559	8949630	
ENSG00000275151	ENST00000614589	CHR_HSCHR5_2_CTG1	29113759	29114317	29113759	29114317	29113759	
ENSG00000275151	ENST00000614589	CHR_HSCHR5_2_CTG1	29113759	29114317	29113759	29114317	29113759	
ENSG00000231049	ENST00000435337	11	5560967	5561917	5560967	5561917	5560967	
ENSG00000280296	ENST00000624531	13	52497097	52500195	52497097	52500195	52500195	
ENSG00000229336	ENST00000282964	21	19739709	19740150	19739709	19740150	19739709	
ENSG00000275008	ENST00000617404	CHR_HSCHR19KIR_GRC212_AB_HAP_CTG3_1	54738018	54752516	54738018	54752516	54738018	54738018
ENSG00000275008	ENST00000617404	CHR_HSCHR19KIR_GRC212_AB_HAP_CTG3_1	54738018	54752516	54738018	54752516	54738018	
ENSG00000275008	ENST00000617404	CHR_HSCHR19KIR_GRC212_AB_HAP_CTG3_1	54738018	54752516	54738018	54752516	54738018	

Datasets -> Filters (filtering and inputs) -> Attributes (desired output) -> Results

Biomart tutorial: [YouTube](#) | [YouKu](#)

Human genome overview ... x AmiGO 2: Welcome x Ontology Lookup Service (... x http://www.e...97c9816967f4 x

www.ensembl.org/biomart/martview/39fe1f97d1569ca10c6c97c9816967f4

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

Search all species...

New Count Results URL XML Perl Help

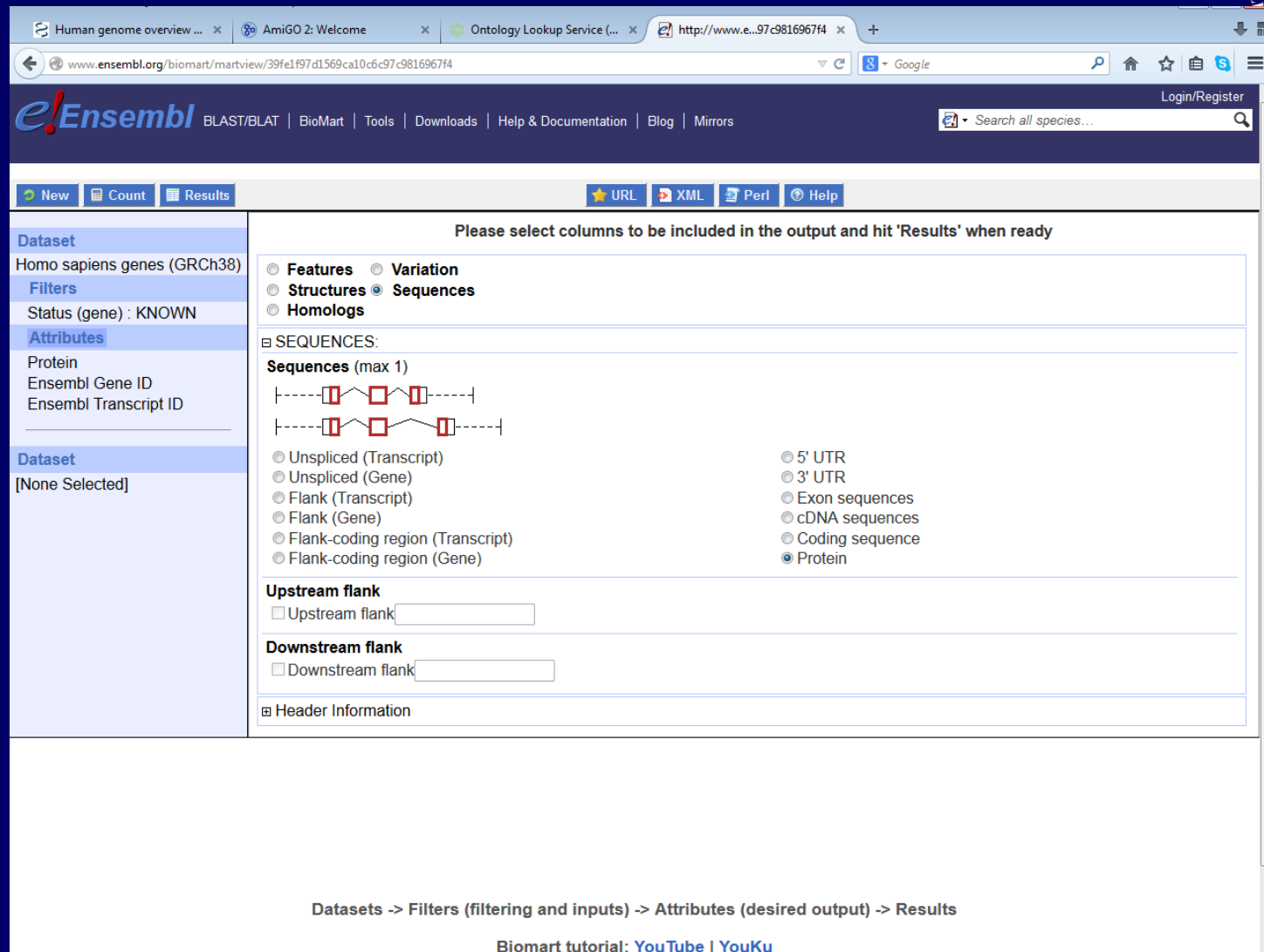
Dataset
Homo sapiens genes (GRCh38)
Filters
Status (gene) : KNOWN
Attributes
Ensembl Gene ID
Ensembl Transcript ID
Chromosome Name
Gene Start (bp)
Gene End (bp)
Transcript Start (bp)
Transcript End (bp)
Transcription Start Site (TSS)
5' UTR Start
5' UTR End
3' UTR Start
3' UTR End
Description
Gene Biotype

Export all results to File TSV Unique results only Go
Email notification to
View 10 rows as HTML Unique XLS

Ensembl Gene ID	Ensembl Transcript ID	Chromosome Name	Gene Start (bp)	Gene End (bp)	Transcript Start (bp)	Transcript End (bp)	Transcription Start Site (TSS)	5' UTR Start
ENSG00000276385	ENST00000618935	20	8934984	8935481	8934984	8935481	8935481	
ENSG00000197468	ENST00000508957	4	8949630	8950559	8949630	8950559	8949630	
ENSG00000275151	ENST00000614589	CHR_HSCHR5_2_CTG1	29113759	29114317	29113759	29114317	29113759	
ENSG00000275151	ENST00000614589	CHR_HSCHR5_2_CTG1	29113759	29114317	29113759	29114317	29113759	
ENSG00000231049	ENST00000435337	11	5560967	5561917	5560967	5561917	5560967	
ENSG00000280296	ENST00000624531	13	52497097	52500195	52497097	52500195	52500195	
ENSG00000229336	ENST00000282964	21	19739709	19740150	19739709	19740150	19739709	
ENSG00000275008	ENST00000617404	CHR_HSCHR19KIR_GRC212_AB_HAP_CTG3_1	54738018	54752516	54738018	54752516	54738018	54738018
ENSG00000275008	ENST00000617404	CHR_HSCHR19KIR_GRC212_AB_HAP_CTG3_1	54738018	54752516	54738018	54752516	54738018	
ENSG00000275008	ENST00000617404	CHR_HSCHR19KIR_GRC212_AB_HAP_CTG3_1	54738018	54752516	54738018	54752516	54738018	

Datasets -> Filters (filtering and inputs) -> Attributes (desired output) -> Results

Biomart tutorial: [YouTube](#) | [YouKu](#)



Human genome overview ... x AmiGO 2: Welcome x Ontology Lookup Service (... x http://www.e...97c9816967f4 x +

www.ensembl.org/biomart/martview/39fe1f97d1569ca10c6c97c9816967f4

Google

Login/Register

Search all species...

New Count Results URL XML Perl Help

Dataset

Homo sapiens genes (GRCh38)

Filters

Status (gene) : KNOWN

Attributes

Protein

Ensembl Gene ID

Ensembl Transcript ID

Dataset

[None Selected]

Please select columns to be included in the output and hit 'Results' when ready

☐ Features ☐ Variation

☐ Structures ☒ Sequences

☐ Homologs

☐ SEQUENCES:

Sequences (max 1)

-----[]-----

-----[]-----

☐ Unspliced (Transcript)

☐ Unspliced (Gene)

☐ Flank (Transcript)

☐ Flank (Gene)

☐ Flank-coding region (Transcript)

☐ Flank-coding region (Gene)

☐ 5' UTR

☐ 3' UTR

☐ Exon sequences

☐ cDNA sequences

☐ Coding sequence

☒ Protein

Upstream flank

☐ Upstream flank

Downstream flank

☐ Downstream flank

☐ Header Information

Datasets -> Filters (filtering and inputs) -> Attributes (desired output) -> Results

Biomart tutorial: [YouTube](#) | [YouKu](#)

Human genome overview ... x AmiGO 2: Welcome x Ontology Lookup Service (... x http://www.e...97c9816967f4 x +

www.ensembl.org/biomart/martview/39fe1f97d1569ca10c6c97c9816967f4

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

Login/Register

Search all species...

New Count Results URL XML Perl Help

Dataset

Homo sapiens genes (GRCh38)

Filters

Status (gene) : KNOWN

Attributes

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

Dataset

[None Selected]

Please select columns to be included in the output and hit 'Results' when ready

☒ **Features** ☐ **Variation**
☐ **Structures** ☐ **Sequences**
☐ **Homologs**

☒ **GENE:**

☒ **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

Human genome overview ... x AmiGO 2: Welcome x Ontology Lookup Service (... x http://www.e...97c9816967f4 x +

www.ensembl.org/biomart/martview/39fe1f97d1569ca10c6c97c9816967f4

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

Login/Register

Search all species...

New Count Results URL XML Perl Help

Dataset
Homo sapiens genes (GRCh38)
Filters
Status (gene) : KNOWN
Attributes
Ensembl Gene ID
Ensembl Transcript ID
Variation Name

Dataset
[None Selected]

Please select columns to be included in the output and hit 'Results' when ready

☐ Features ☒ Variation
☐ Structures ☐ Sequences
☐ Homologs

☐ GENE:

☐ GERMLINE VARIATION INFORMATION:

Variation

☒ Variation Name ☐ Mapweight
☐ Variation Source ☐ Minor allele
☐ Variation source description ☐ Minor allele frequency
☐ Variant Alleles ☐ Minor allele count
☐ Evidence status ☐ Clinical significance

Variation Location

☐ Transcript location (bp) ☐ Protein location (aa)
☐ Variation Chromosome Strand ☐ Chromosome Location (bp)

Gene Location and Effect

☐ PolyPhen prediction ☐ CDS Start
☐ PolyPhen score ☐ CDS End
☐ SIFT prediction ☐ Protein Allele
☐ SIFT score ☐ Consequence Type (Transcript Variation)
☐ Distance to transcript ☐ Consequence specific allele

☐ SOMATIC VARIATION INFORMATION:

Datasets -> Filters (filtering and inputs) -> Attributes (desired output) -> Results

Biomart tutorial: [YouTube](#) | [YouKu](#)

Human genome overview ... x AmiGO 2: Welcome x Ontology Lookup Service (... x http://www.e...97c9816967f4 x +

www.ensembl.org/biomart/martview/39fe1f97d1569ca10c6c97c9816967f4

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

Login/Register

Search all species...

New Count Results URL XML Perl Help

Dataset

Homo sapiens genes (GRCh38)

Filters

Status (gene) : KNOWN

Attributes

Ensembl Gene ID

Ensembl Transcript ID

Dataset

[None Selected]

Please select columns to be included in the output and hit 'Results' when ready

☐ Features ☐ Variation

☐ Structures ☐ Sequences

☒ Homologs

☐ GENE:

☐ ORTHOLOGS (Max select 6 orthologs):

Alpaca Orthologs

☐ Alpaca Ensembl Gene ID ☐ Ancestor

☐ Canonical Protein or Transcript ID ☐ Orthology confidence [0 low, 1 high]

☐ Alpaca Ensembl Protein ID ☐ % Identity with respect to query gene

☐ Alpaca Chromosome Name ☐ % Identity with respect to Alpaca gene

☐ Alpaca Chromosome Start (bp) ☐ dN

☐ Alpaca Chromosome End (bp) ☐ dS

☐ Homology Type

Amazon molly Orthologs

☐ Amazon molly Ensembl Gene ID ☐ Homology Type

☐ Canonical Protein or Transcript ID ☐ Ancestor

☐ Amazon molly Ensembl Protein ID ☐ Orthology confidence [0 low, 1 high]

☐ Amazon molly Chromosome Name ☐ % Identity with respect to query gene

☐ Amazon molly Chromosome Start (bp) ☐ % Identity with respect to Amazon molly gene

☐ Amazon molly Chromosome End (bp)

Anole Lizard Orthologs

☐ Anole Lizard Ensembl Gene ID ☐ Ancestor

☐ Canonical Protein or Transcript ID ☐ Orthology confidence [0 low, 1 high]

☐ Anole Lizard Ensembl Protein ID ☐ % Identity with respect to query gene

☐ Anole Lizard Chromosome Name ☐ % Identity with respect to Anole Lizard gene

☐ Anole Lizard Chromosome Start (bp) ☐ dN

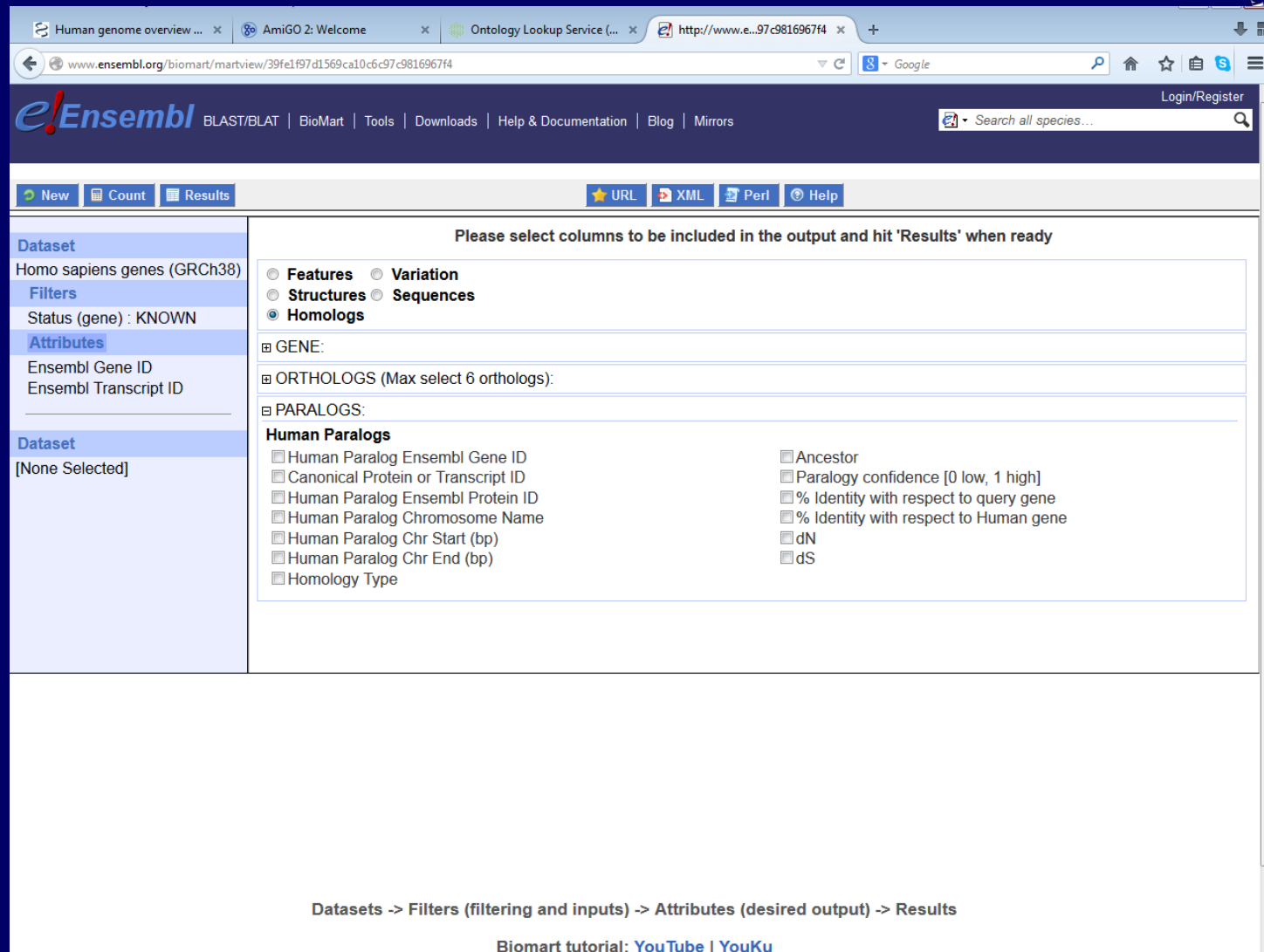
☐ Anole Lizard Chromosome End (bp) ☐ dS

☐ Homology Type

Armadillo Orthologs

☐ Armadillo Ensembl Gene ID ☐ Ancestor

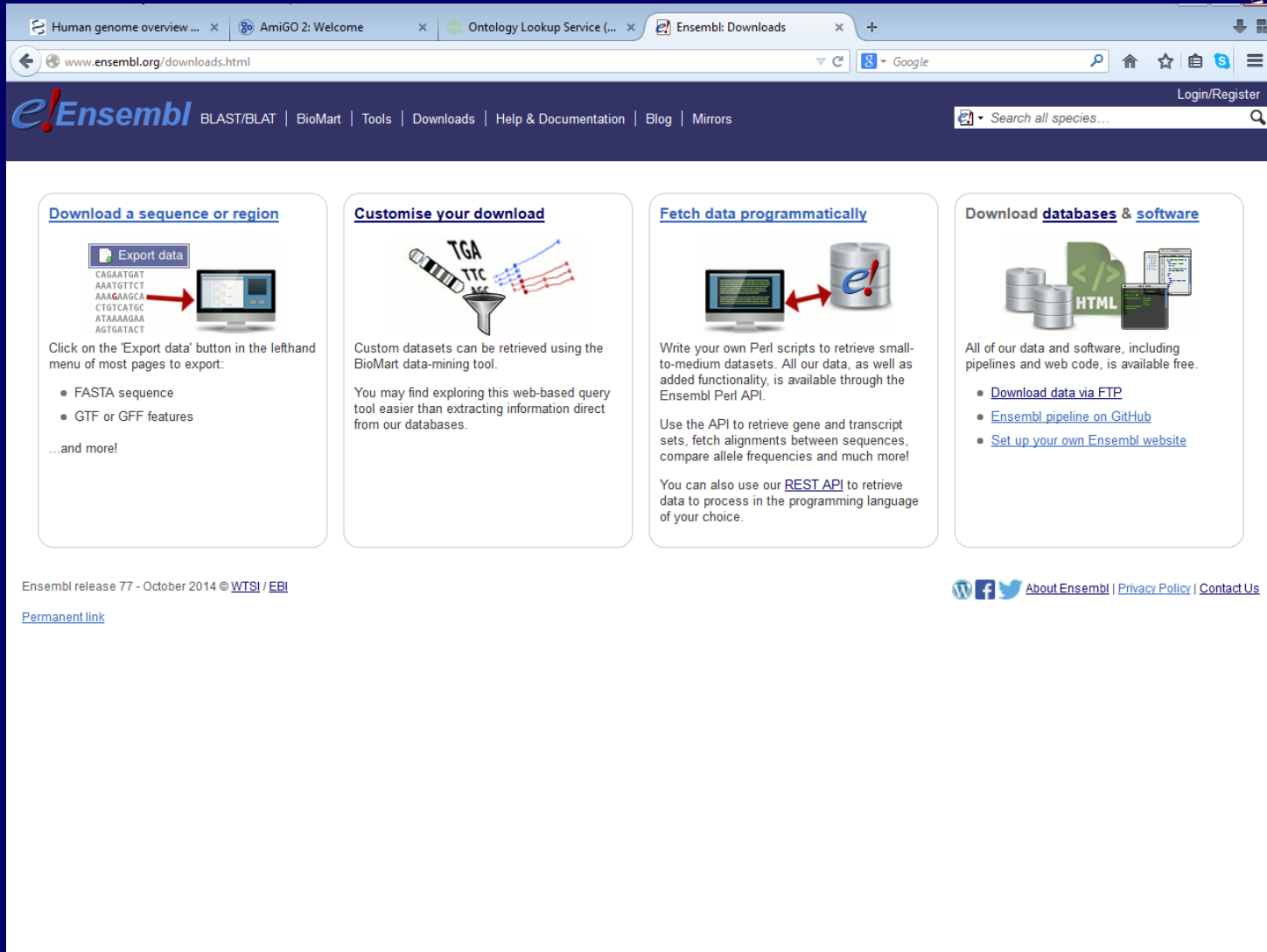
☐ Canonical Protein or Transcript ID ☐ Orthology confidence [0 low, 1 high]



The screenshot shows the Ensembl Biomart web interface. The browser tabs include 'Human genome overview ...', 'AmiGO 2: Welcome', 'Ontology Lookup Service (...)', and 'http://www.e...97c9816967f4'. The address bar shows 'www.ensembl.org/biomart/martview/39fe1f97d1569ca10c6c97c9816967f4'. The Ensembl logo is at the top left, with navigation links for BLAST/BLAT, BioMart, Tools, Downloads, Help & Documentation, Blog, and Mirrors. A search bar is on the top right with the text 'Search all species...'. Below the navigation bar, there are buttons for 'New', 'Count', and 'Results'. The main content area is titled 'Please select columns to be included in the output and hit 'Results' when ready'. It contains several sections: 'Dataset' (Homo sapiens genes (GRCh38)), 'Filters' (Status (gene) : KNOWN), 'Attributes' (Ensembl Gene ID, Ensembl Transcript ID), and 'Human Paralogs' (Human Paralog Ensembl Gene ID, Canonical Protein or Transcript ID, Human Paralog Ensembl Protein ID, Human Paralog Chromosome Name, Human Paralog Chr Start (bp), Human Paralog Chr End (bp), Homology Type, Ancestor, Paralogy confidence [0 low, 1 high], % Identity with respect to query gene, % Identity with respect to Human gene, dN, dS). At the bottom, there is a navigation bar with the text 'Datasets -> Filters (filtering and inputs) -> Attributes (desired output) -> Results' and a link to 'Biomart tutorial: YouTube | YouKu'.

A teljes adatbázis letöltése

- FTP (File Transfer Protocol)
- Az összes faj összes adata ami az ENSEMBL-ben van
 - Szekvencia adatok
 - Teljes változatban
 - Részben feldolgozva
 - Annotációs adatok



Human genome overview ... x AmiGO 2: Welcome x Ontology Lookup Service (... x Ensembl: Downloads x

www.ensembl.org/downloads.html

Google

Login/Register

Search all species...

Download a sequence or region

Export data

CAGAAATGAT
AAATGTCT
AAGAAGCA
CTGTCAATGC
ATAAAGAA
AGTGATCT

Click on the 'Export data' button in the lefthand menu of most pages to export:

- FASTA sequence
- GTF or GFF features

...and more!

Customise your download

TGA
TTC
ATC

Custom datasets can be retrieved using the BioMart data-mining tool.

You may find exploring this web-based query tool easier than extracting information direct from our databases.

Fetch data programmatically

Write your own Perl scripts to retrieve small-to-medium datasets. All our data, as well as added functionality, is available through the Ensembl Perl API.

Use the API to retrieve gene and transcript sets, fetch alignments between sequences, compare allele frequencies and much more!

You can also use our [REST API](#) to retrieve data to process in the programming language of your choice.

Download databases & software

All of our data and software, including pipelines and web code, is available free.

- [Download data via FTP](#)
- [Ensembl pipeline on GitHub](#)
- [Set up your own Ensembl website](#)

Ensembl release 77 - October 2014 © WTSI / EBI

[Permanent link](#)

[About Ensembl](#) | [Privacy Policy](#) | [Contact Us](#)

Human genome overview ... x AmiGO 2: Welcome x Ontology Lookup Service (... x FTP Download x

www.ensembl.org/info/data/ftp/index.html

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

Using this website | Annotation and prediction | Data access | API & software | About us

In this section
Downloading with rsync

Search documentation:

FTP Download

You can download via a browser from our [FTP site](#), use a script, or even use [rsync](#) from the command line.

API Code

If you do not have access to git, you can obtain our latest API code as a gzipped tarball:
[Download complete API for this release](#)

Note: the API version needs to be the same as the databases you are accessing, so please use git to obtain a previous version if querying older databases.

Database dumps

Entire databases can be downloaded from our FTP site in a variety of formats. Please be aware that some of these files can run to many gigabytes of data.

Looking for [MySQL dumps](#) to install databases locally? See our [web installation instructions](#) for full details.

Each directory on [ftp.ensembl.org](#) contains a [README](#) file, explaining the directory structure.

Multi-species data

Database	MySQL	EMF	MAF	BED	XML	Ancestral Alleles
Comparative genomics	MySQL					
BioMart	MySQL	-	-	-	-	-
Stable ids	MySQL	-	-	-	-	-

Single species data

Popular species are listed first. You can customise this list via our [home page](#).

Show	10	entries	Show/hide columns	Filter										
★	Species	DNA (FASTA)	cDNA (FASTA)	CDS (FASTA)	ncRNA (FASTA)	Protein sequence (FASTA)	Annotated sequence (EMBL)	Annotated sequence (GenBank)	Gene sets	Whole databases	Variation (GVF)	Variation (VCF)	Variation (VEP)	Regulation (GFF)
Y	Human <i>Homo sapiens</i>	FASTA	FASTA	FASTA	FASTA	FASTA	EMBL	GenBank	GTF	MySQL	GVF	VCF	VEP	Regulation (GFF)
Y	Mouse <i>Mus musculus</i>	FASTA	FASTA	FASTA	FASTA	FASTA	EMBL	GenBank	GTF	MySQL	GVF	VCF	VEP	Regulation (GFF)
Y	Zebrafish <i>Danio rerio</i>	FASTA	FASTA	FASTA	FASTA	FASTA	EMBL	GenBank	GTF	MySQL	GVF	VCF	VEP	-

Human genome overview ... x AmiGO 2: Welcome x Ontology Lookup Service (... x FTP Download x Index of ftp://ftp.ensembl.org/pub/release-77/fasta/homo_sapiens/dna/ x +

ftp://ftp.ensembl.org/pub/release-77/fasta/homo_sapiens/dna/ Google

Index of ftp://ftp.ensembl.org/pub/release-77/fasta/homo_sapiens/dna/

 Up to higher level directory

Name	Size	Last Modified
 CHECKSUMS	63 KB	9/23/2014 7:14:00 PM
 Homo_sapiens.GRCh38.dna.chromosome.10.fa.gz	39002 KB	9/23/2014 10:46:00 AM
 Homo_sapiens.GRCh38.dna.chromosome.11.fa.gz	39079 KB	9/23/2014 10:46:00 AM
 Homo_sapiens.GRCh38.dna.chromosome.12.fa.gz	38774 KB	9/23/2014 10:46:00 AM
 Homo_sapiens.GRCh38.dna.chromosome.13.fa.gz	28827 KB	9/23/2014 10:46:00 AM
 Homo_sapiens.GRCh38.dna.chromosome.14.fa.gz	26520 KB	9/23/2014 10:46:00 AM
 Homo_sapiens.GRCh38.dna.chromosome.15.fa.gz	24457 KB	9/23/2014 10:46:00 AM
 Homo_sapiens.GRCh38.dna.chromosome.16.fa.gz	23566 KB	9/23/2014 10:46:00 AM
 Homo_sapiens.GRCh38.dna.chromosome.17.fa.gz	23334 KB	9/23/2014 10:46:00 AM
 Homo_sapiens.GRCh38.dna.chromosome.18.fa.gz	22471 KB	9/23/2014 10:46:00 AM
 Homo_sapiens.GRCh38.dna.chromosome.19.fa.gz	16294 KB	9/23/2014 10:46:00 AM
 Homo_sapiens.GRCh38.dna.chromosome.1.fa.gz	67627 KB	9/23/2014 10:46:00 AM
 Homo_sapiens.GRCh38.dna.chromosome.20.fa.gz	18271 KB	9/23/2014 10:46:00 AM
 Homo_sapiens.GRCh38.dna.chromosome.21.fa.gz	11505 KB	9/23/2014 10:46:00 AM
 Homo_sapiens.GRCh38.dna.chromosome.22.fa.gz	11115 KB	9/23/2014 10:46:00 AM
 Homo_sapiens.GRCh38.dna.chromosome.2.fa.gz	71031 KB	9/23/2014 10:46:00 AM
 Homo_sapiens.GRCh38.dna.chromosome.3.fa.gz	58223 KB	9/23/2014 10:46:00 AM
 Homo_sapiens.GRCh38.dna.chromosome.4.fa.gz	55900 KB	9/23/2014 10:46:00 AM
 Homo_sapiens.GRCh38.dna.chromosome.5.fa.gz	53171 KB	9/23/2014 10:46:00 AM
 Homo_sapiens.GRCh38.dna.chromosome.6.fa.gz	50223 KB	9/23/2014 10:46:00 AM
 Homo_sapiens.GRCh38.dna.chromosome.7.fa.gz	46184 KB	9/23/2014 10:46:00 AM
 Homo_sapiens.GRCh38.dna.chromosome.8.fa.gz	42495 KB	9/23/2014 10:46:00 AM
 Homo_sapiens.GRCh38.dna.chromosome.9.fa.gz	35599 KB	9/23/2014 10:46:00 AM
 Homo_sapiens.GRCh38.dna.chromosome.CHR_HG142_HG150_NOVEL_TEST.fa.gz	455 KB	9/23/2014 10:46:00 AM
 Homo_sapiens.GRCh38.dna.chromosome.CHR_HG151_NOVEL_TEST.fa.gz	451 KB	9/23/2014 10:46:00 AM
 Homo_sapiens.GRCh38.dna.chromosome.CHR_HSCR10_1_CTG1.fa.gz	445 KB	9/23/2014 10:46:00 AM
 Homo_sapiens.GRCh38.dna.chromosome.CHR_HSCR10_1_CTG2.fa.gz	484 KB	9/23/2014 10:46:00 AM
 Homo_sapiens.GRCh38.dna.chromosome.CHR_HSCR10_1_CTG3.fa.gz	441 KB	9/23/2014 10:46:00 AM
 Homo_sapiens.GRCh38.dna.chromosome.CHR_HSCR10_1_CTG4.fa.gz	450 KB	9/23/2014 10:46:00 AM
 Homo_sapiens.GRCh38.dna.chromosome.CHR_HSCR11_1_CTG1_1.fa.gz	440 KB	9/23/2014 10:46:00 AM
 Homo_sapiens.GRCh38.dna.chromosome.CHR_HSCR11_1_CTG2.fa.gz	450 KB	9/23/2014 10:46:00 AM
 Homo_sapiens.GRCh38.dna.chromosome.CHR_HSCR11_1_CTG3.fa.gz	414 KB	9/23/2014 10:46:00 AM

```

File Edit View History Bookmarks Tools Help
Human genome overview - Genome... x The Gene Ontology x Ontology Lookup Service (OLS) x FTP Download x ftp://ftp.ensembl.org/piens/dna/README x
ftp://ftp.ensembl.org/pub/release-73/fasta/homo_sapiens/dna/README

#### README ####

IMPORTANT: Please note you can download correlation data tables,
supported by Ensembl, via the highly customisable BioMart and
Ensembl data mining tools. See http://www.ensembl.org/biomart/martview or
http://www.ebi.ac.uk/biomart/ for more information.

The genome assembly represented here corresponds to GenBank Assembly ID
GCA_000001405.13

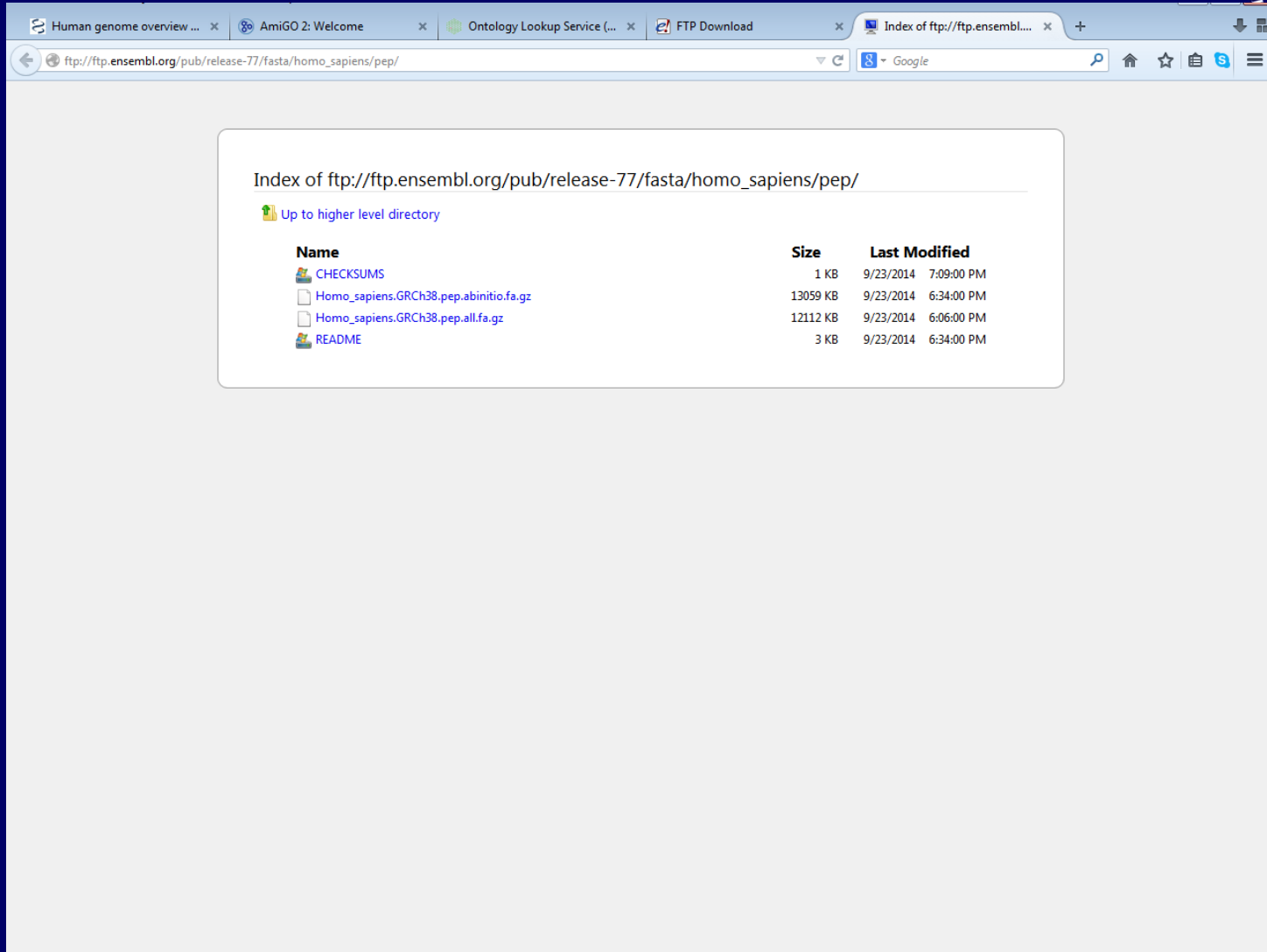
#####
Fasta DNA dumps
#####

-----
FILE NAMES
-----
The files are consistently named following this pattern:
<species>.<assembly>.<release>.<sequence type>.<id type>.<id>.fa.gz

<species>: The systematic name of the species.
<assembly>: The assembly build name.
<release>: The release number.
<sequence type>:
* 'dna' - unmasked genomic DNA sequences.
* 'dna_rm' - masked genomic DNA. Interspersed repeats and low
complexity regions are detected with the RepeatMasker tool and masked
by replacing repeats with 'N's.
* 'dna_sm' - soft-masked genomic DNA. All repeats and low complexity regions
have been replaced with lowercased versions of their nucleic base
<id type> One of the following:
* 'chromosome' - The top-level coordinate system in most species in Ensembl
* 'nonchromosomal' - Contains DNA that has not been assigned a chromosome
* 'seqlevel' - This is usually sequence scaffolds, chunks or clones.
-- 'scaffold' - Larger sequence contigs from the assembly of shorter
sequencing reads (often from whole genome shotgun, WGS) which could
not yet be assembled into chromosomes. Often more genome sequencing
is needed to narrow gaps and establish a tiling path.
-- 'chunk' - While contig sequences can be assembled into large entities,
they sometimes have to be artificially broken down into smaller entities
called 'chunks'. This is due to limitations in the annotation
pipeline and the finite record size imposed by MySQL which stores the
sequence and annotation information.
-- 'clone' - In general this is the smallest sequence entity. It is often
identical to the sequence of one BAC clone, or sequence region
of one BAC clone which forms the tiling path.
<id>: The actual sequence identifier. Depending on the <id type> the <id>
could represent the name of a chromosome, a scaffold, a contig, a clone ..
Field is empty for seqlevel files
fa: All files in these directories represent FASTA database files
gz: All files are compacted with GNU Zip for storage efficiency.

EXAMPLES
The genomic sequence of human chromosome 1:
Homo_sapiens.GRCh37.57.dna.chromosome.1.fa.gz
The masked version of the genome sequence of human chromosome 1:

```



Human genome overview ... x AmiGO 2: Welcome x Ontology Lookup Service (... x FTP Download x Index of ftp://ftp.ensembl.org/pub/release-77/fasta/homo_sapiens/pep/

ftp://ftp.ensembl.org/pub/release-77/fasta/homo_sapiens/pep/ Google

Index of ftp://ftp.ensembl.org/pub/release-77/fasta/homo_sapiens/pep/

[Up to higher level directory](#)

Name	Size	Last Modified
CHECKSUMS	1 KB	9/23/2014 7:09:00 PM
Homo_sapiens.GRCh38.pep.abinitio.fa.gz	13059 KB	9/23/2014 6:34:00 PM
Homo_sapiens.GRCh38.pep.all.fa.gz	12112 KB	9/23/2014 6:06:00 PM
README	3 KB	9/23/2014 6:34:00 PM

Human genome overview ... x AmiGO 2: Welcome x Ontology Lookup Service (... x FTP Download x Index of ftp://ftp.ensembl.org/pub/release-77/fasta/ x +

ftp://ftp.ensembl.org/pub/release-77/fasta/ Google

Index of ftp://ftp.ensembl.org/pub/release-77/fasta/

Up to higher level directory

Name	Size	Last Modified
ailuropoda_melanoleuca		9/28/2014 10:02:00 PM
anas_platyrhynchos		9/28/2014 10:02:00 PM
ancestral_alleles		9/29/2014 5:37:00 AM
anolis_carolinensis		9/28/2014 10:02:00 PM
astyanax_mexicanus		9/28/2014 10:02:00 PM
bos_taurus		9/28/2014 10:02:00 PM
caenorhabditis_elegans		9/28/2014 10:02:00 PM
callithrix_jacchus		9/28/2014 10:02:00 PM
canis_familiaris		9/28/2014 10:02:00 PM
cavia_porcellus		9/28/2014 10:02:00 PM
chlorocebus_sabaeus		9/28/2014 10:02:00 PM
choleopus_hoffmanni		9/28/2014 10:02:00 PM
ciona_intestinalis		9/28/2014 10:02:00 PM
ciona_savignyi		9/28/2014 10:02:00 PM
danio_rerio		9/28/2014 10:02:00 PM
dasyatis_novemcinctus		9/28/2014 10:02:00 PM
dipodomys_ordii		9/28/2014 10:02:00 PM
drosophila_melanogaster		9/28/2014 10:02:00 PM
echinops_telfairi		9/28/2014 10:02:00 PM
equus_caballus		9/28/2014 10:02:00 PM
erinaceus_europaeus		9/28/2014 10:02:00 PM
felis_catus		9/28/2014 10:02:00 PM
ficedula_albicollis		9/28/2014 10:02:00 PM
gadus_morhua		9/28/2014 10:02:00 PM
gallus_gallus		9/28/2014 10:02:00 PM
gasterosteus_aculeatus		9/28/2014 10:02:00 PM
gorilla_gorilla		9/28/2014 10:02:00 PM
homo_sapiens		9/28/2014 10:32:00 PM
ictidomys_tridecemlineatus		9/28/2014 10:32:00 PM

Human genome overview ... x AmiGO 2: Welcome x Ontology Lookup Service (... x FTP Download x Index of ftp://ftp.ensembl.org/pub/release-77/mysql/homo_sapiens_core_77_38/

ftp://ftp.ensembl.org/pub/release-77/mysql/homo_sapiens_core_77_38/ Google

Index of ftp://ftp.ensembl.org/pub/release-77/mysql/homo_sapiens_core_77_38/

Up to higher level directory

Name	Size	Last Modified
alt_allele_attrrib.txt.gz	8 KB	9/30/2014 4:10:00 PM
alt_allele_group.txt.gz	2 KB	9/30/2014 4:10:00 PM
alt_allele.txt.gz	23 KB	9/30/2014 4:10:00 PM
analysis_description.txt.gz	6 KB	9/30/2014 4:10:00 PM
analysis.txt.gz	3 KB	9/30/2014 4:10:00 PM
assembly_exception.txt.gz	6 KB	9/30/2014 4:10:00 PM
assembly.txt.gz	10604 KB	9/30/2014 4:10:00 PM
associated_group.txt.gz	1 KB	9/30/2014 4:10:00 PM
associated_xref.txt.gz	1 KB	9/30/2014 4:10:00 PM
attrib_type.txt.gz	8 KB	9/30/2014 4:10:00 PM
CHECKSUMS.gz	1 KB	9/30/2014 4:18:00 PM
coord_system.txt.gz	1 KB	9/30/2014 4:10:00 PM
data_file.txt.gz	1 KB	9/30/2014 4:10:00 PM
density_feature.txt.gz	267 KB	9/30/2014 4:10:00 PM
density_type.txt.gz	1 KB	9/30/2014 4:10:00 PM
dependent_xref.txt.gz	17218 KB	9/30/2014 4:10:00 PM
ditag_feature.txt.gz	1 KB	9/30/2014 4:10:00 PM
ditag.txt.gz	1 KB	9/30/2014 4:10:00 PM
dna_align_feature.txt.gz	497184 KB	9/30/2014 4:15:00 PM
dna.txt.gz	979008 KB	9/30/2014 4:11:00 PM
exon_transcript.txt.gz	5942 KB	9/30/2014 4:15:00 PM
exon.txt.gz	16263 KB	9/30/2014 4:15:00 PM
external_db.txt.gz	13 KB	9/30/2014 4:15:00 PM
external_synonym.txt.gz	1057 KB	9/30/2014 4:15:00 PM
gene_archive.txt.gz	5234 KB	9/30/2014 4:15:00 PM
gene_attrrib.txt.gz	1317 KB	9/30/2014 4:15:00 PM
gene.txt.gz	3023 KB	9/30/2014 4:15:00 PM
genome_statistics.txt.gz	1 KB	9/30/2014 4:15:00 PM
homo_sapiens_core_77_38.sql.gz	6 KB	9/30/2014 4:18:00 PM
identity_xref.txt.gz	3416 KB	9/30/2014 4:15:00 PM
interpro.txt.gz	143 KB	9/30/2014 4:15:00 PM
intron_supporting_evidence.txt.gz	1 KB	9/30/2014 4:15:00 PM

Mit tanultunk ma?

- Az integrált adatbázisok nagyon hatékony eszközök
- Rugalmasak és könnyen használhatók
- Az adatbázisok egymástól függenek – ez veszélyes lehet
- A tökéletes adatbázis a bioinformatika végső célja



Feladat 6.

- Keresd meg a neked legérdekesebb fehérjét vagy gént valamelyik ENSEMBL genomban és próbálj megtudni minnél többet róla.