Bioinformatika és genomanalízis az orvostudományban

Integrált biológiai adatbázisok

Cserző Miklós

2020

https://semmelweis.zoom.us/j/96102872458?pwd=Rk1PL2tqS21sdlUwc3B4eDFCZkNKQT09

A mai előadás

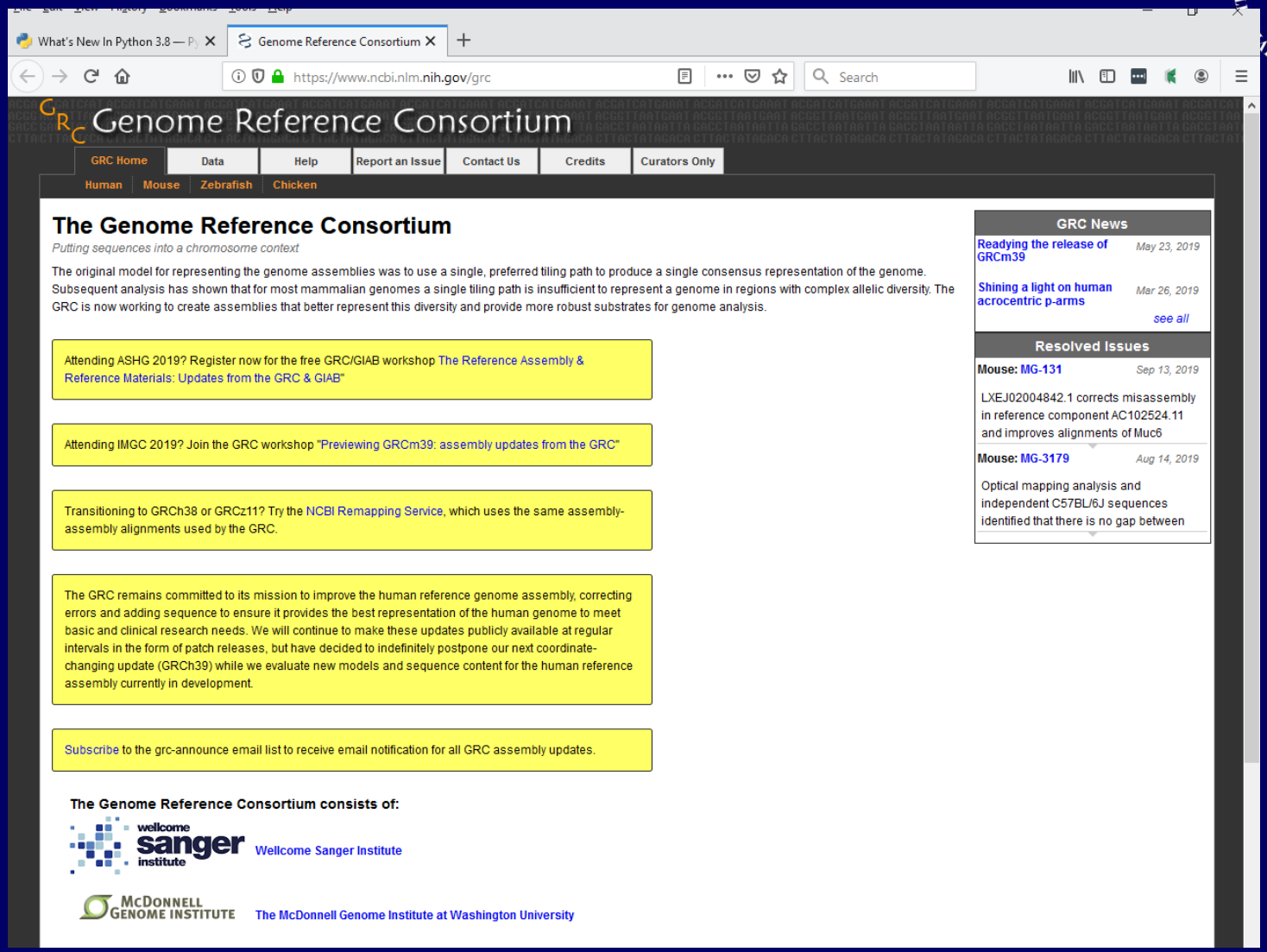
- A genom annotálás jelentősége
- Genome Reference Consortium
- Gene Ontology
- Az „ensembl” pipeline
- Lekérdezés a web-felületen keresztül
- Az osztott annotációs rendszer (DAS)
- A BioMart felület használata
- Adatelérés FTP-portálon

Miért annotálunk genomot?

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


What's New In Python 3.8
Genome Reference Consortium
https://www.ncbi.nlm.nih.gov/grc
Search
Genome Reference Consortium
GRC Home
Data
Help
Report an Issue
Contact Us
Credits
Curators Only
Human
Mouse
Zebrafish
Chicken
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 2019? Register now for the free GRC/GIAB workshop The Reference Assembly & Reference Materials: Updates from the GRC & GIAB"
Attending IMGC 2019? Join the GRC workshop "Previewing GRCm39: assembly updates from the GRC"
Transitioning to GRCh38 or GRCz11? Try the NCBI Remapping Service, which uses the same assembly-assembly alignments used by the GRC.
The GRC remains committed to its mission to improve the human reference genome assembly, correcting errors and adding sequence to ensure it provides the best representation of the human genome to meet basic and clinical research needs. We will continue to make these updates publicly available at regular intervals in the form of patch releases, but have decided to indefinitely postpone our next coordinate-changing update (GRCh39) while we evaluate new models and sequence content for the human reference assembly currently in development.
Subscribe to the grc-announce email list to receive email notification for all GRC assembly updates.
The Genome Reference Consortium consists of:
wellcome sanger institute
Wellcome Sanger Institute
McDonnell Genome Institute
The McDonnell Genome Institute at Washington University
GRC News
Readying the release of GRCm39
May 23, 2019
Shining a light on human acrocentric p-arms
Mar 26, 2019
see all
Resolved Issues
Mouse: MG-131
Sep 13, 2019
LXEJ02004842.1 corrects misassembly in reference component AC102524.11 and improves alignments of Muc6
Mouse: MG-3179
Aug 14, 2019
Optical mapping analysis and independent C57BL/6J sequences identified that there is no gap between

What's New In Python 3.8 — P | Genome Assemblies - Genome

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

Search

Genome Reference Consortium

GRC Home | Data | Help | Report an Issue | Contact Us | Credits | Curators Only

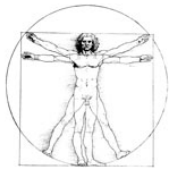
Human | Mouse | Zebrafish | Chicken | Paper Supplemental Data

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.

GRC News

Readying the release of GRCm39	May 23, 2019
Shining a light on human acrocentric p-arms	Mar 26, 2019
see all	



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	p13
Patches released	FIX: 113, NOVEL: 72

More human assembly statistics...

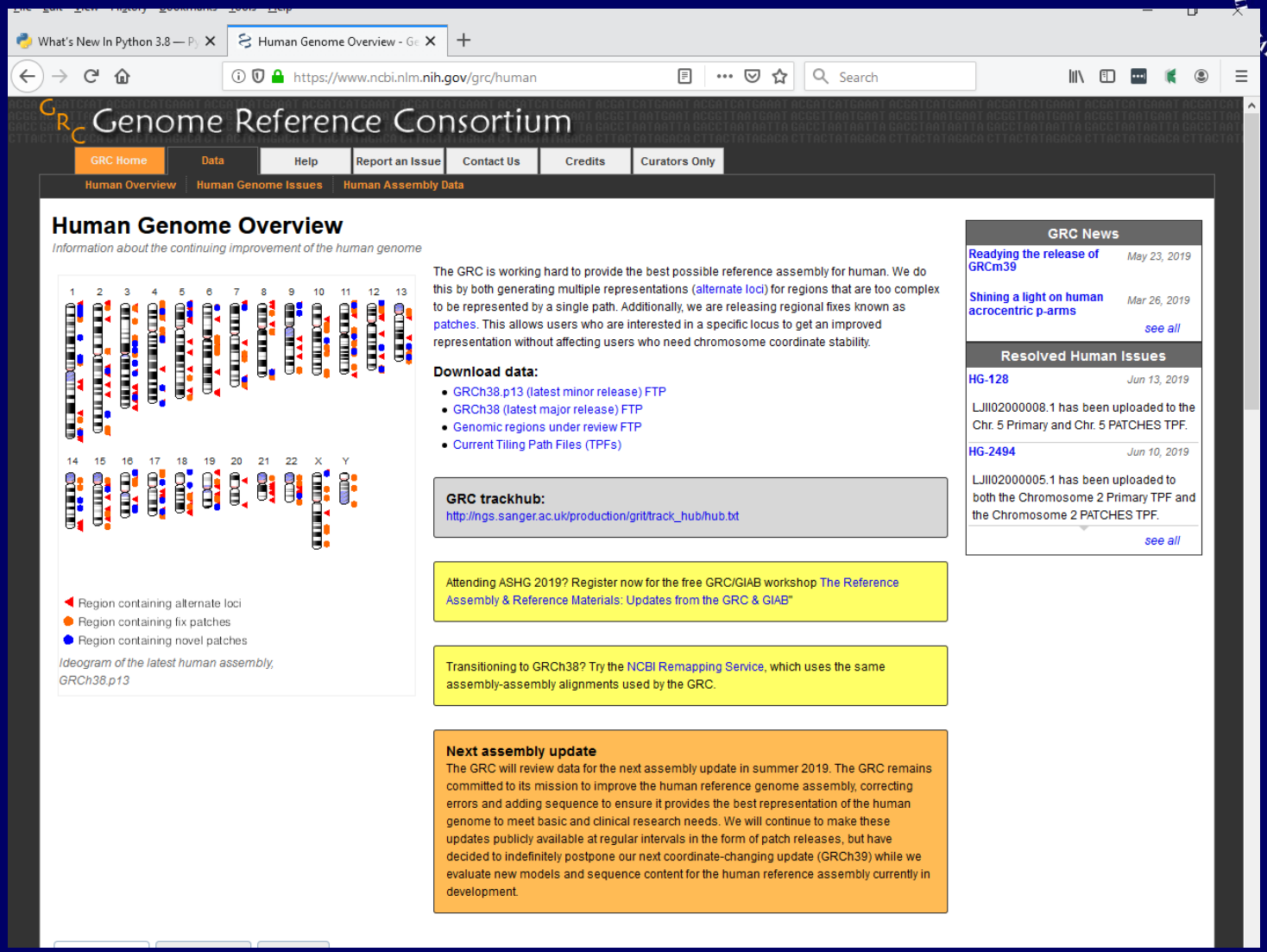
Mouse

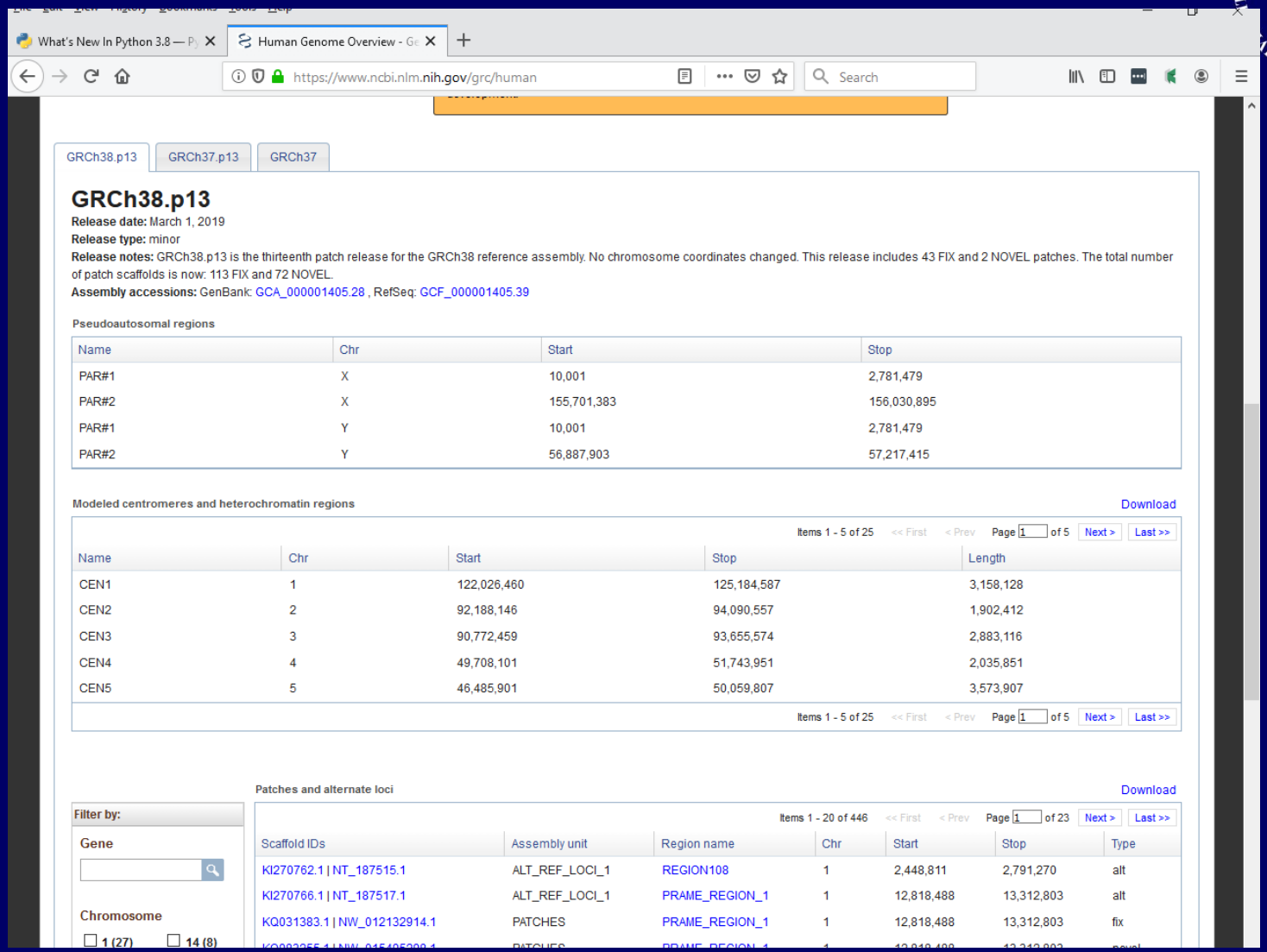
The GRC has produced an updated assembly (GRCm38). This is an update of the last MGSC assembly (MGSCv37) which was described in 2009 (PMID: 19468303). The primary assembly is based on assembling overlapping BAC clones derived from the C57BL/6J strain and several loci have sequence available from other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

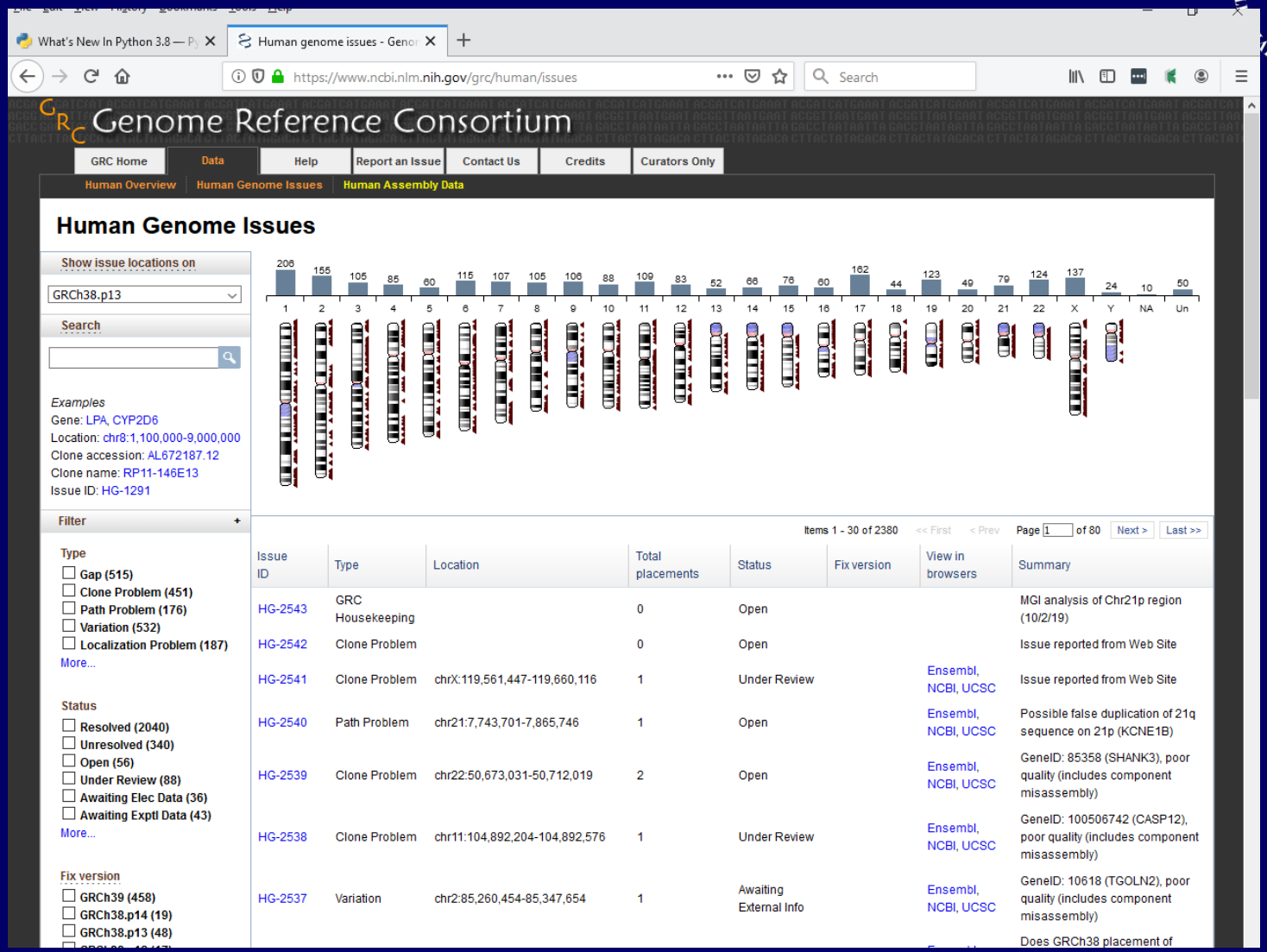
Mouse assembly information

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

More mouse assembly statistics...


What's New In Python 3.8
Human Genome Overview
https://www.ncbi.nlm.nih.gov/grc/human
Search
Genome Reference Consortium
GRC Home
Data
Help
Report an Issue
Contact Us
Credits
Curators Only
Human Overview
Human Genome Issues
Human Assembly Data
Human Genome Overview
Information about the continuing improvement of the human genome
1 2 3 4 5 6 7 8 9 10 11 12 13
14 15 16 17 18 19 20 21 22 X Y
Region containing alternate loci
Region containing fix patches
Region containing novel patches
Ideogram of the latest human assembly, GRCh38.p13
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.p13 (latest minor release) FTP
GRCh38 (latest major release) FTP
Genomic regions under review FTP
Current Tiling Path Files (TPFs)
GRC trackhub:
http://ngs.sanger.ac.uk/production/grit/track_hub/hub.txt
Attending ASHG 2019? Register now for the free GRC/GIAB workshop The Reference Assembly & Reference Materials: Updates from the GRC & GIAB"
Transitioning to GRCh38? Try the NCBI Remapping Service, which uses the same assembly-assembly alignments used by the GRC.
Next assembly update
The GRC will review data for the next assembly update in summer 2019. The GRC remains committed to its mission to improve the human reference genome assembly, correcting errors and adding sequence to ensure it provides the best representation of the human genome to meet basic and clinical research needs. We will continue to make these updates publicly available at regular intervals in the form of patch releases, but have decided to indefinitely postpone our next coordinate-changing update (GRCh39) while we evaluate new models and sequence content for the human reference assembly currently in development.
GRC News
Readying the release of GRCm39 May 23, 2019
Shining a light on human acrocentric p-arms Mar 26, 2019
see all
Resolved Human Issues
HG-128 Jun 13, 2019
LJII02000008.1 has been uploaded to the Chr. 5 Primary and Chr. 5 PATCHES TPF.
HG-2494 Jun 10, 2019
LJII02000005.1 has been uploaded to both the Chromosome 2 Primary TPF and the Chromosome 2 PATCHES TPF.
see all


What's New In Python 3.8
Human Genome Overview
https://www.ncbi.nlm.nih.gov/grc/human
Search
GRCh38.p13
GRCh37.p13
GRCh37
GRCh38.p13
Release date: March 1, 2019
Release type: minor
Release notes: GRCh38.p13 is the thirteenth patch release for the GRCh38 reference assembly. No chromosome coordinates changed. This release includes 43 FIX and 2 NOVEL patches. The total number of patch scaffolds is now: 113 FIX and 72 NOVEL.
Assembly accessions: GenBank: GCA_000001405.28 , RefSeq: GCF_000001405.39
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
Chromosome
1 (27)
14 (8)
Items 1 - 20 of 446 << First < Prev Page 1 of 23 Next > Last >>
Scaffold IDs Assembly unit Region name Chr Start Stop Type
KI270762.1 | NT_187515.1 ALT_REF_LOCI_1 REGION108 1 2,448,811 2,791,270 alt
KI270766.1 | NT_187517.1 ALT_REF_LOCI_1 PRAME_REGION_1 1 12,818,488 13,312,803 alt
KQ031383.1 | NW_012132914.1 PATCHES PRAME_REGION_1 1 12,818,488 13,312,803 fix


What's New In Python 3.8
Human genome issues - Genor
https://www.ncbi.nlm.nih.gov/grc/human/issues
Search
Genome Reference Consortium
GRC Home
Data
Help
Report an Issue
Contact Us
Credits
Curators Only
Human Overview
Human Genome Issues
Human Assembly Data
Human Genome Issues
Show issue locations on
GRCh38.p13
Search
Examples
Gene: LPA, CYP2D6
Location: chr8:1,100,000-9,000,000
Clone accession: AL672187.12
Clone name: RP11-146E13
Issue ID: HG-1291
Filter
Type
Gap (515)
Clone Problem (451)
Path Problem (176)
Variation (532)
Localization Problem (187)
More...
Status
Resolved (2040)
Unresolved (340)
Open (56)
Under Review (88)
Awaiting Elec Data (36)
Awaiting Exptl Data (43)
More...
Fix version
GRCh39 (458)
GRCh38.p14 (19)
GRCh38.p13 (48)
206 155 105 85 60 115 107 105 106 88 109 83 52 66 76 60 162 44 123 49 79 124 137 24 10 50
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 X Y NA Un
Items 1 - 30 of 2380
<< First
< Prev
Page 1 of 80
Next >
Last >>
Issue ID
Type
Location
Total placements
Status
Fix version
View in browsers
Summary
HG-2543 GRC Housekeeping 0 Open MGI analysis of Chr21p region (10/2/19)
HG-2542 Clone Problem 0 Open Issue reported from Web Site
HG-2541 Clone Problem chrX:119,561,447-119,660,116 1 Under Review Ensembl, NCBI, UCSC Issue reported from Web Site
HG-2540 Path Problem chr21:7,743,701-7,865,746 1 Open Ensembl, NCBI, UCSC Possible false duplication of 21q sequence on 21p (KCNE1B)
HG-2539 Clone Problem chr22:50,673,031-50,712,019 2 Open Ensembl, NCBI, UCSC GeneID: 85358 (SHANK3), poor quality (includes component misassembly)
HG-2538 Clone Problem chr11:104,892,204-104,892,576 1 Under Review Ensembl, NCBI, UCSC GeneID: 100506742 (CASP12), poor quality (includes component misassembly)
HG-2537 Variation chr2:85,260,454-85,347,654 1 Awaiting External Info Ensembl, NCBI, UCSC GeneID: 10618 (TGOLN2), poor quality (includes component misassembly)
Does GRCh38 placement of

Mivel annotáljunk?

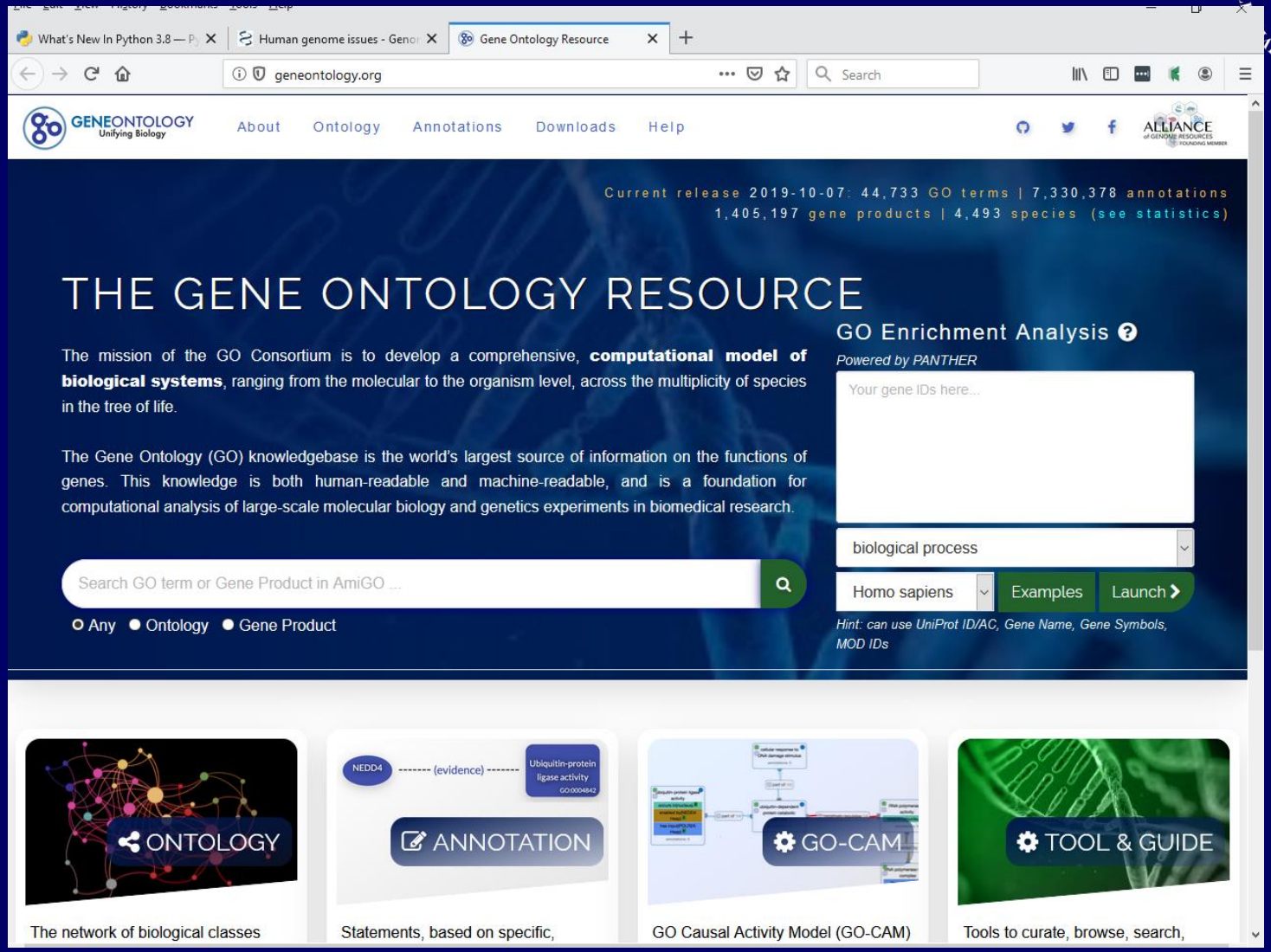
- Gene Ontology Consortium
- Zárt annotációs szótár – egy bizottság dönt, mi kerülhet bele
- Cél: a gének pontos és teljes leírása
- Követelmény: a leírás legyen koherens, általános és gépi feldolgozásra alkalmas

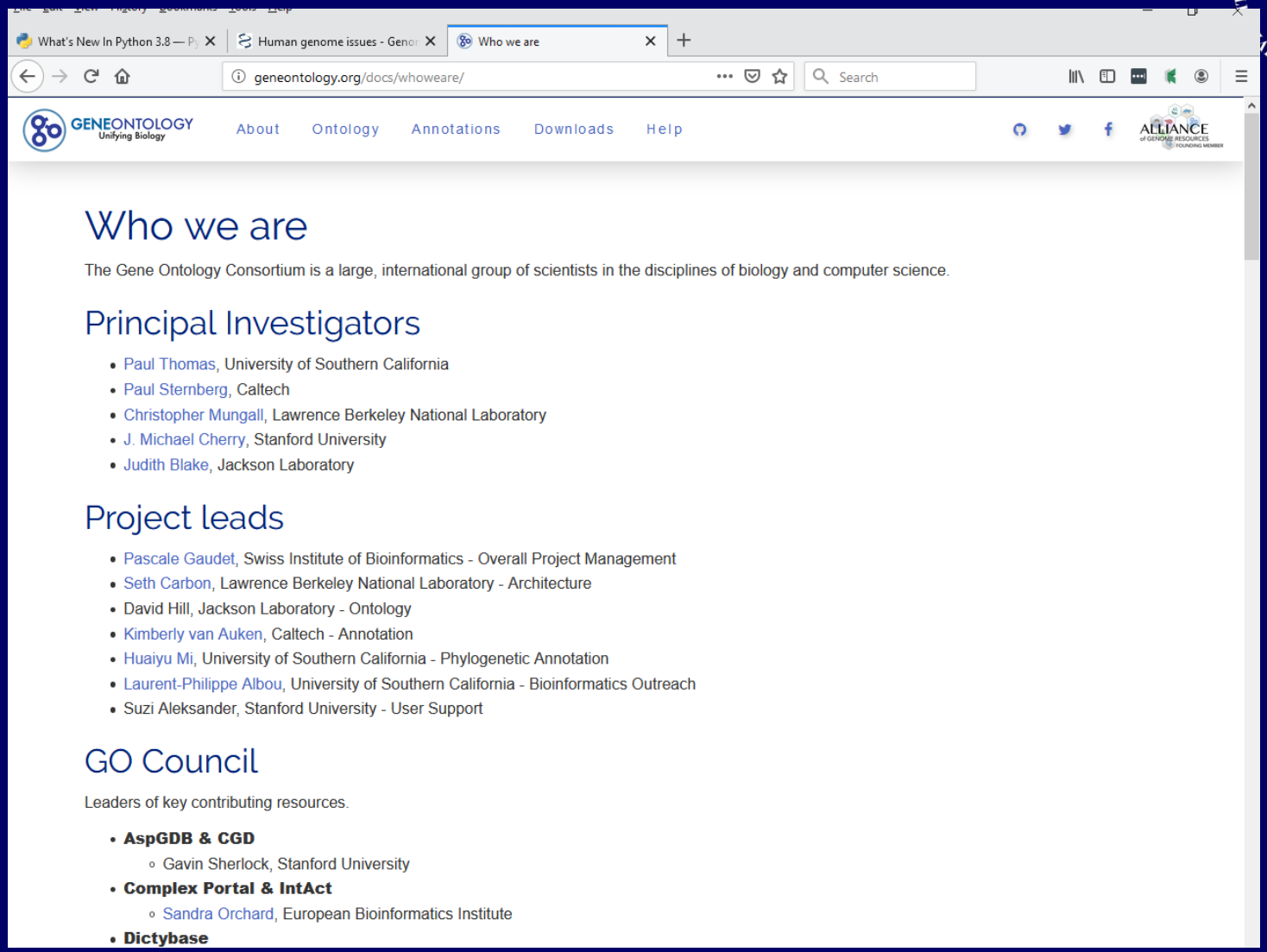
A megvalósítás

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

A módszer

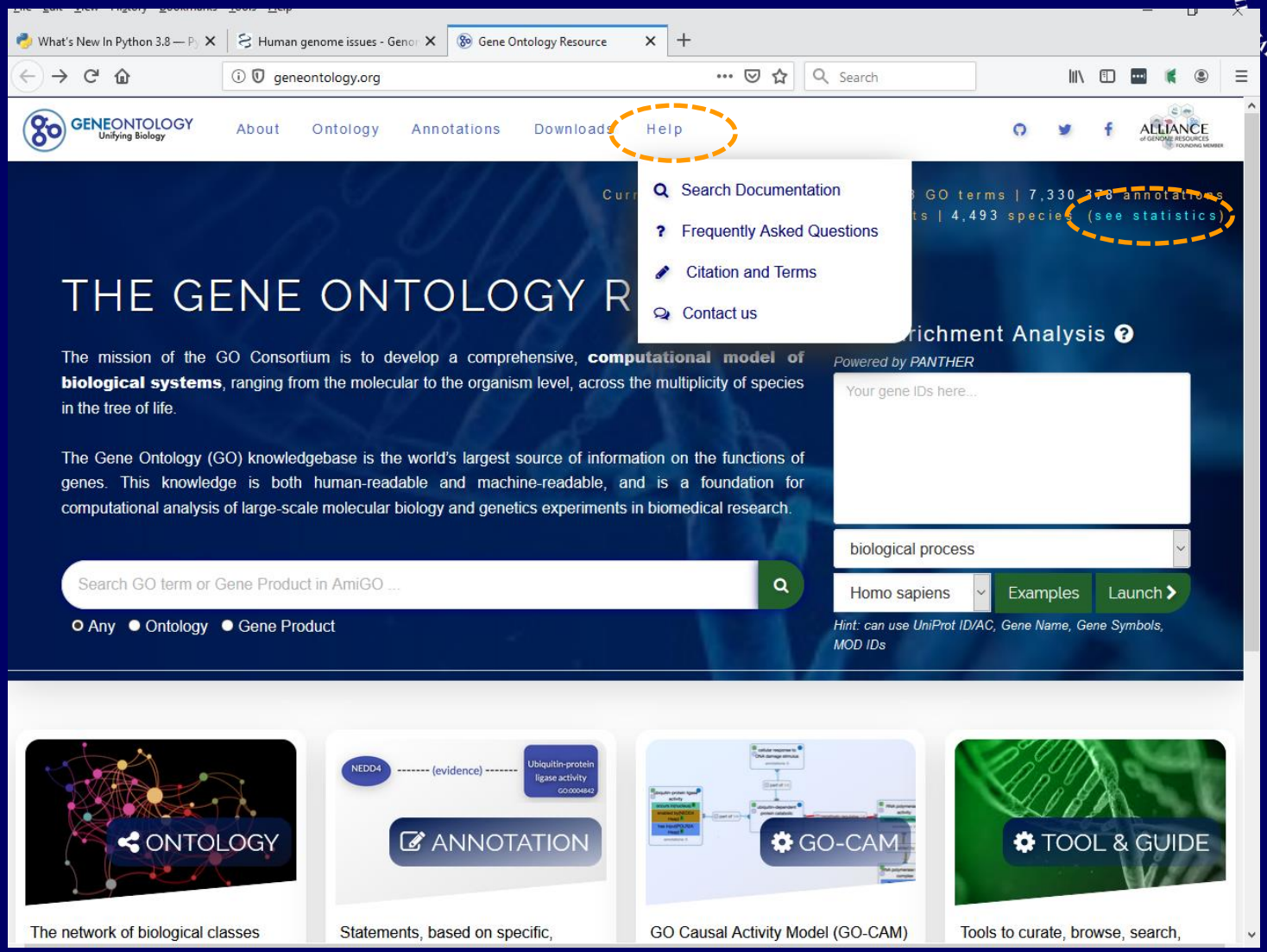
- 12 modell-lény – és egy népes szakértői csapat
- Elsődleges forrás: közvetlen kisérletes adatok
- Másodlagos forrás: homológ fehérjék annotációja
- Azt is jelzik, ha nincs adat
- Honlap: http://www.geneontology.org/

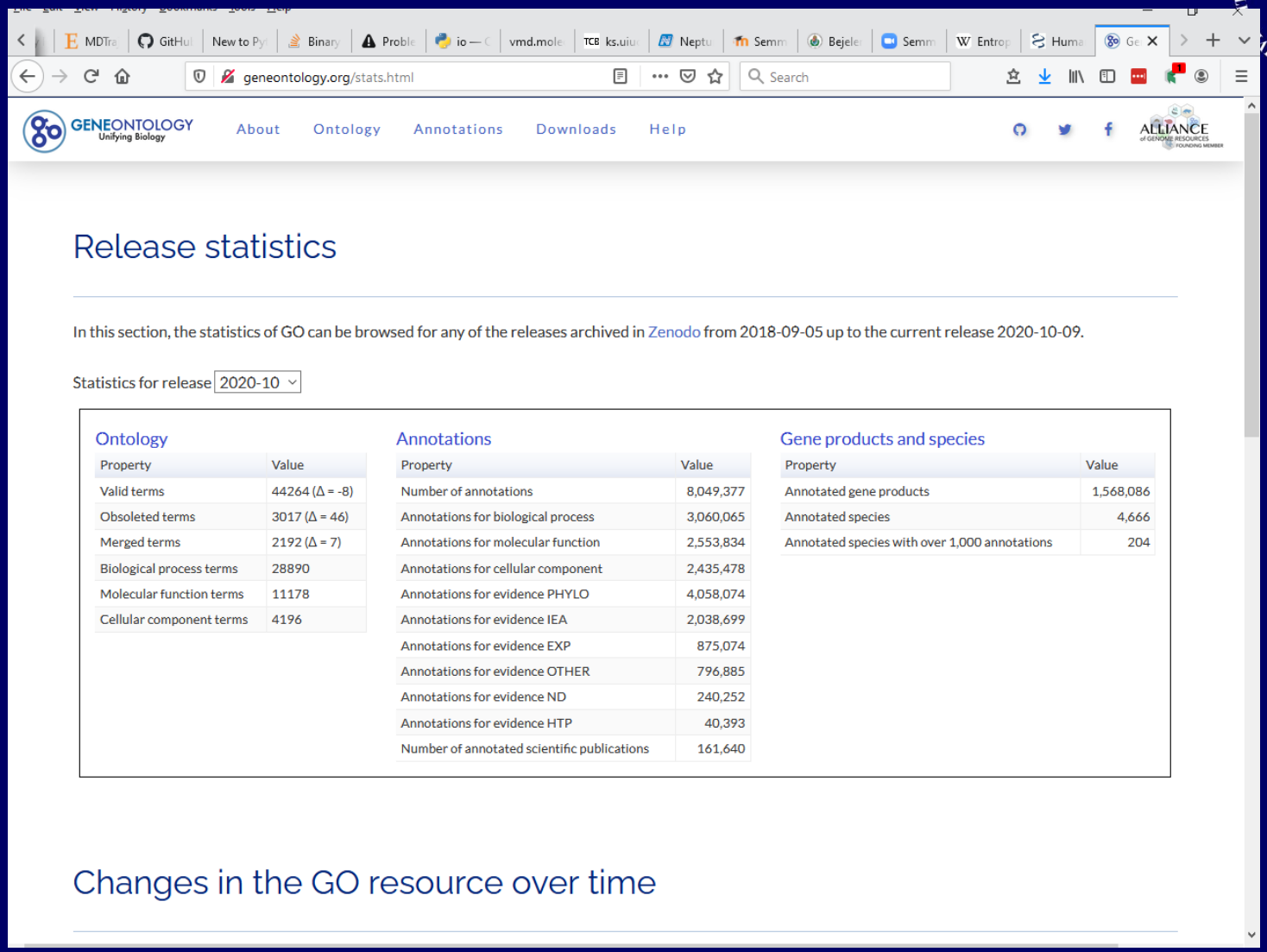

What's New In Python 3.8
Human genome issues - Geno
Gene Ontology Resource
geneontology.org
Search
GENEONTOLOGY
Unifying Biology
About
Ontology
Annotations
Downloads
Help
ALLIANCE
Current release 2019-10-07: 44,733 GO terms | 7,330,378 annotations
1,405,197 gene products | 4,493 species (see statistics)
THE GENE ONTOLOGY RESOURCE
The mission of the GO Consortium is to develop a comprehensive, computational model of biological systems, ranging from the molecular to the organism level, across the multiplicity of species in the tree of life.
The Gene Ontology (GO) knowledgebase is the world's largest source of information on the functions of genes. This knowledge is both human-readable and machine-readable, and is a foundation for computational analysis of large-scale molecular biology and genetics experiments in biomedical research.
Search GO term or Gene Product in AmiGO ...
Any
Ontology
Gene Product
GO Enrichment Analysis
Powered by PANTHER
Your gene IDs here...
biological process
Homo sapiens
Examples
Launch
Hint: can use UniProt ID/AC, Gene Name, Gene Symbols, MOD IDs
ONTOLOGY
The network of biological classes
NEDD4
(evidence)
Ubiquitin-protein ligase activity
GO:0004842
ANNOTATION
Statements, based on specific,
GO-CAM
GO Causal Activity Model (GO-CAM)
TOOL & GUIDE
Tools to curate, browse, search,

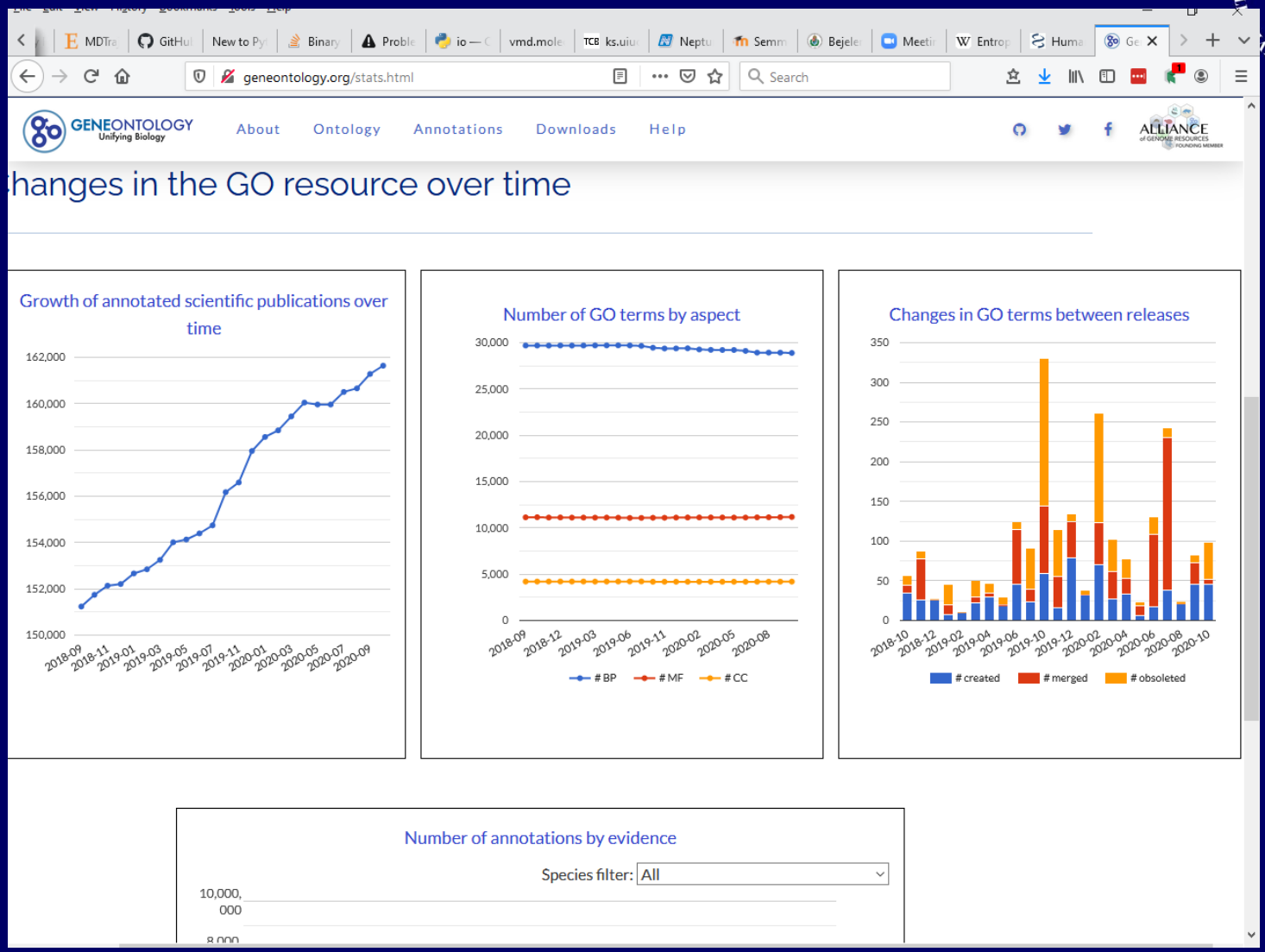

What's New In Python 3.8
Human genome issues - Geno
Who we are
geneontology.org/docs/whoweare/
Search
GENEONTOLOGY
Unifying Biology
About
Ontology
Annotations
Downloads
Help
ALLIANCE
Who we are
The Gene Ontology Consortium is a large, international group of scientists in the disciplines of biology and computer science.
Principal Investigators
Paul Thomas, University of Southern California
Paul Sternberg, Caltech
Christopher Mungall, Lawrence Berkeley National Laboratory
J. Michael Cherry, Stanford University
Judith Blake, Jackson Laboratory
Project leads
Pascale Gaudet, Swiss Institute of Bioinformatics - Overall Project Management
Seth Carbon, Lawrence Berkeley National Laboratory - Architecture
David Hill, Jackson Laboratory - Ontology
Kimberly van Auken, Caltech - Annotation
Huaiyu Mi, University of Southern California - Phylogenetic Annotation
Laurent-Philippe Albou, University of Southern California - Bioinformatics Outreach
Suzi Aleksander, Stanford University - User Support
GO Council
Leaders of key contributing resources.
AspGDB & CGD
Gavin Sherlock, Stanford University
Complex Portal & IntAct
Sandra Orchard, European Bioinformatics Institute
Dictybase

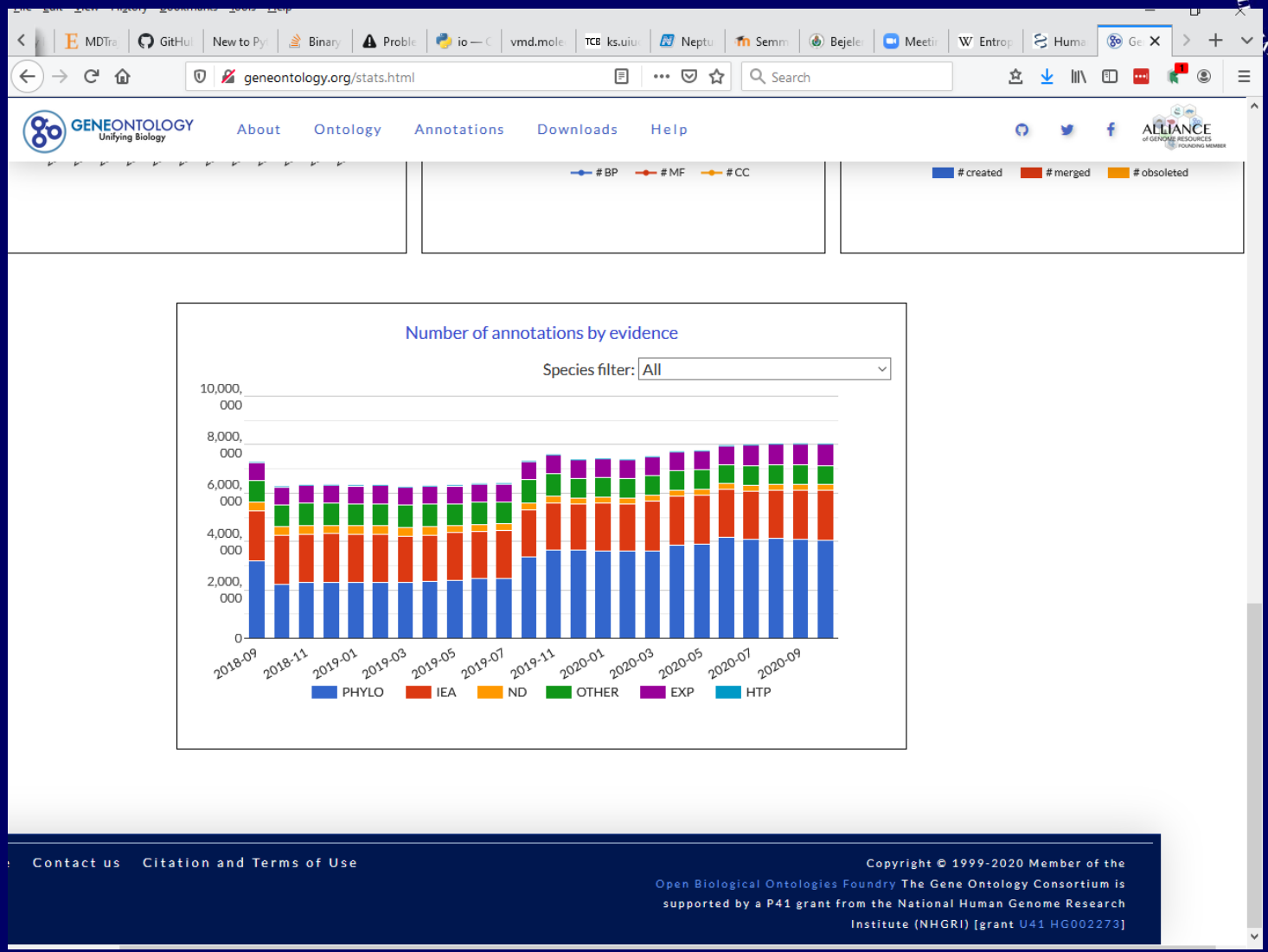
Meeting Photo: GO Consortium Meeting

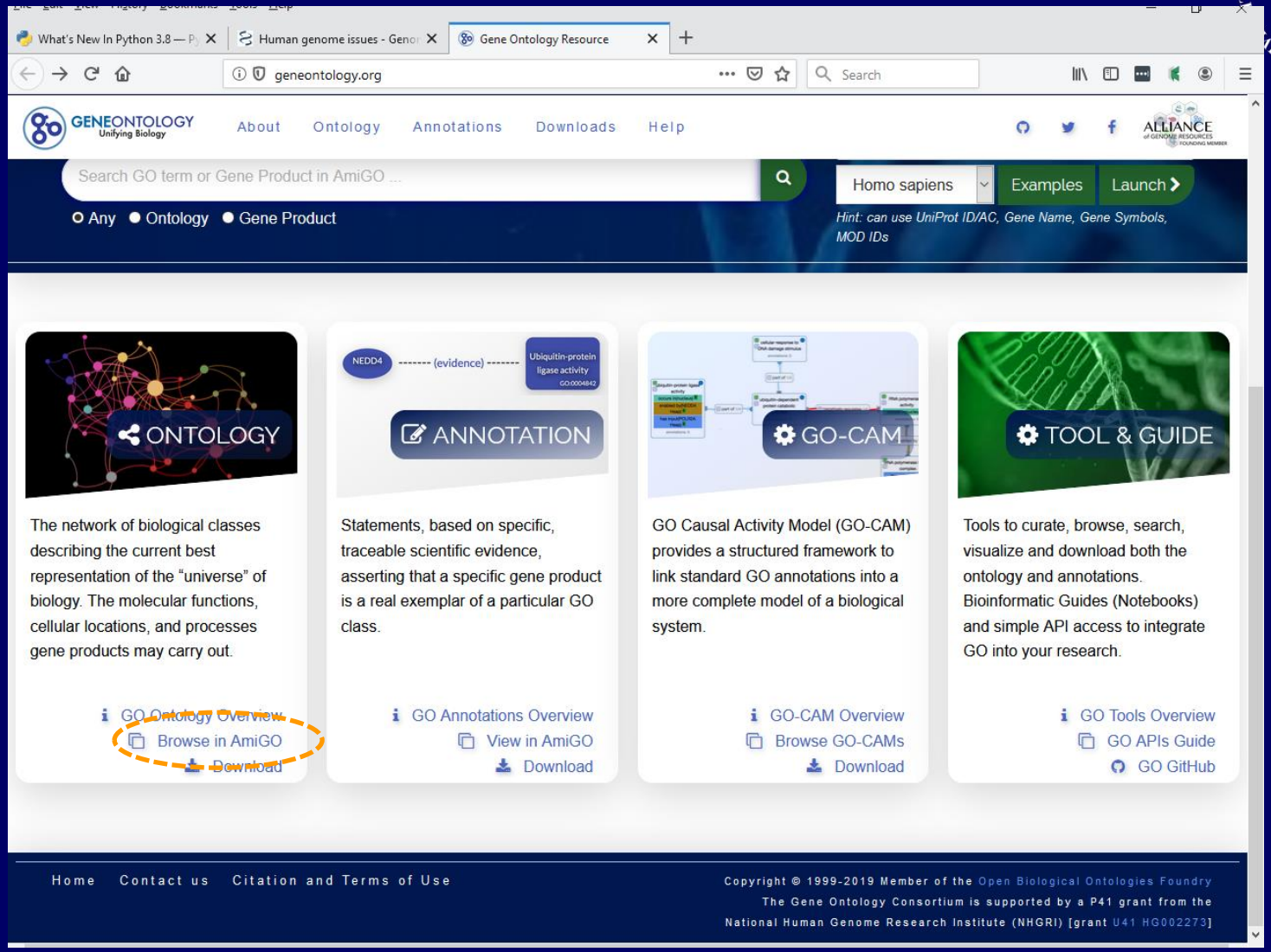
Hôtel Mont Gabriel, Sainte-Adèle, Québec, Canada, October 21-23, 2008

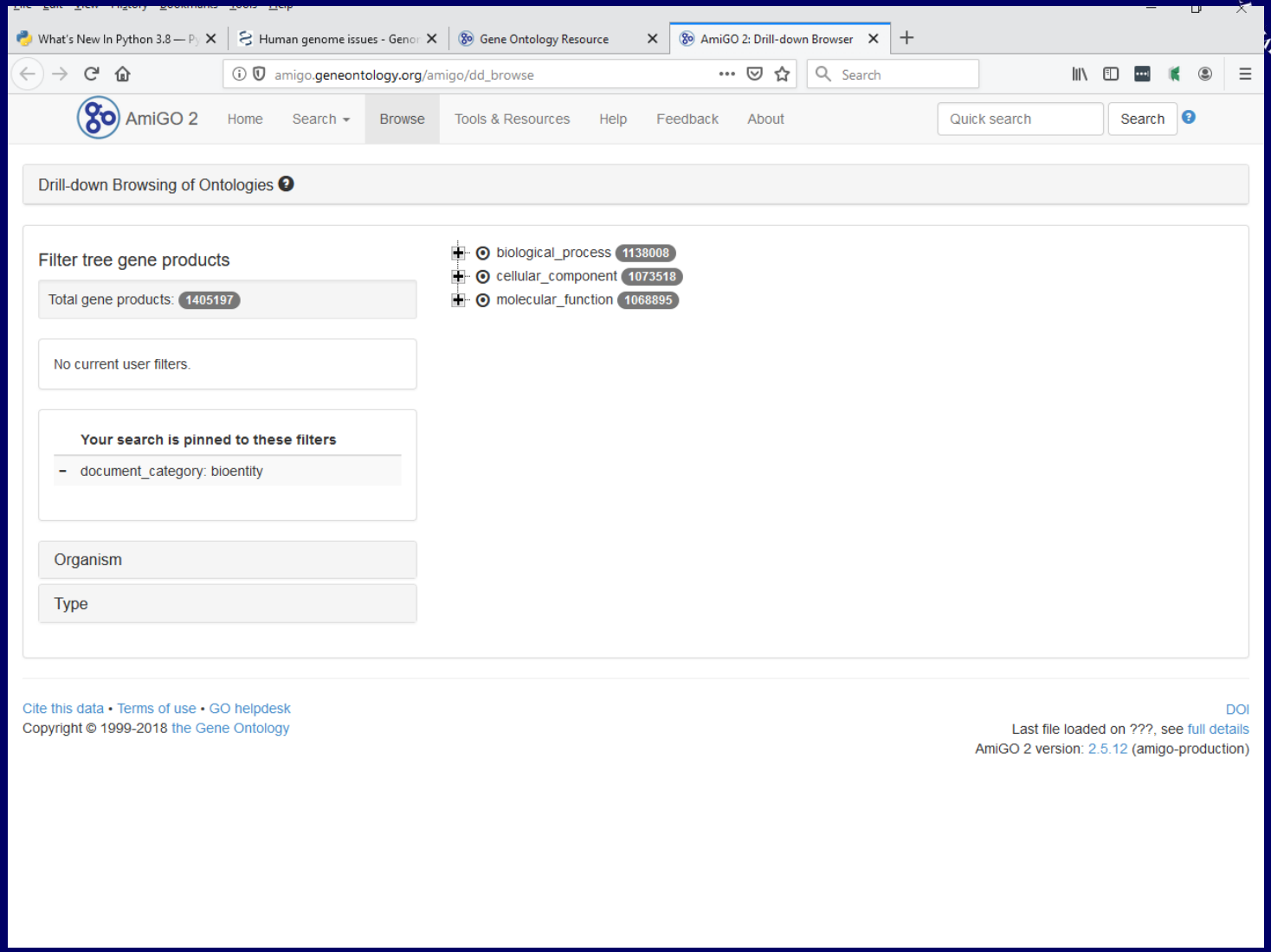

What's New In Python 3.8
Human genome issues - Geno
Gene Ontology Resource
geneontology.org
Search
GENEONTOLOGY
Unifying Biology
About
Ontology
Annotations
Downloads
Help
ALLIANCE
Search Documentation
Frequently Asked Questions
Citation and Terms
Contact us
GO terms | 7,330,378 annotations
ts | 4,493 species
(see statistics)
THE GENE ONTOLOGY R
The mission of the GO Consortium is to develop a comprehensive, computational model of biological systems, ranging from the molecular to the organism level, across the multiplicity of species in the tree of life.
The Gene Ontology (GO) knowledgebase is the world's largest source of information on the functions of genes. This knowledge is both human-readable and machine-readable, and is a foundation for computational analysis of large-scale molecular biology and genetics experiments in biomedical research.
richment Analysis
Powered by PANTHER
Your gene IDs here...
biological process
Homo sapiens
Examples
Launch
Hint: can use UniProt ID/AC, Gene Name, Gene Symbols, MOD IDs
Search GO term or Gene Product in AmiGO ...
Any
Ontology
Gene Product
ONTOLOGY
NEDD4
(evidence)
Ubiquitin-protein ligase activity
GO:0004842
ANNOTATION
GO-CAM
TOOL & GUIDE
The network of biological classes
Statements, based on specific,
GO Causal Activity Model (GO-CAM)
Tools to curate, browse, search,

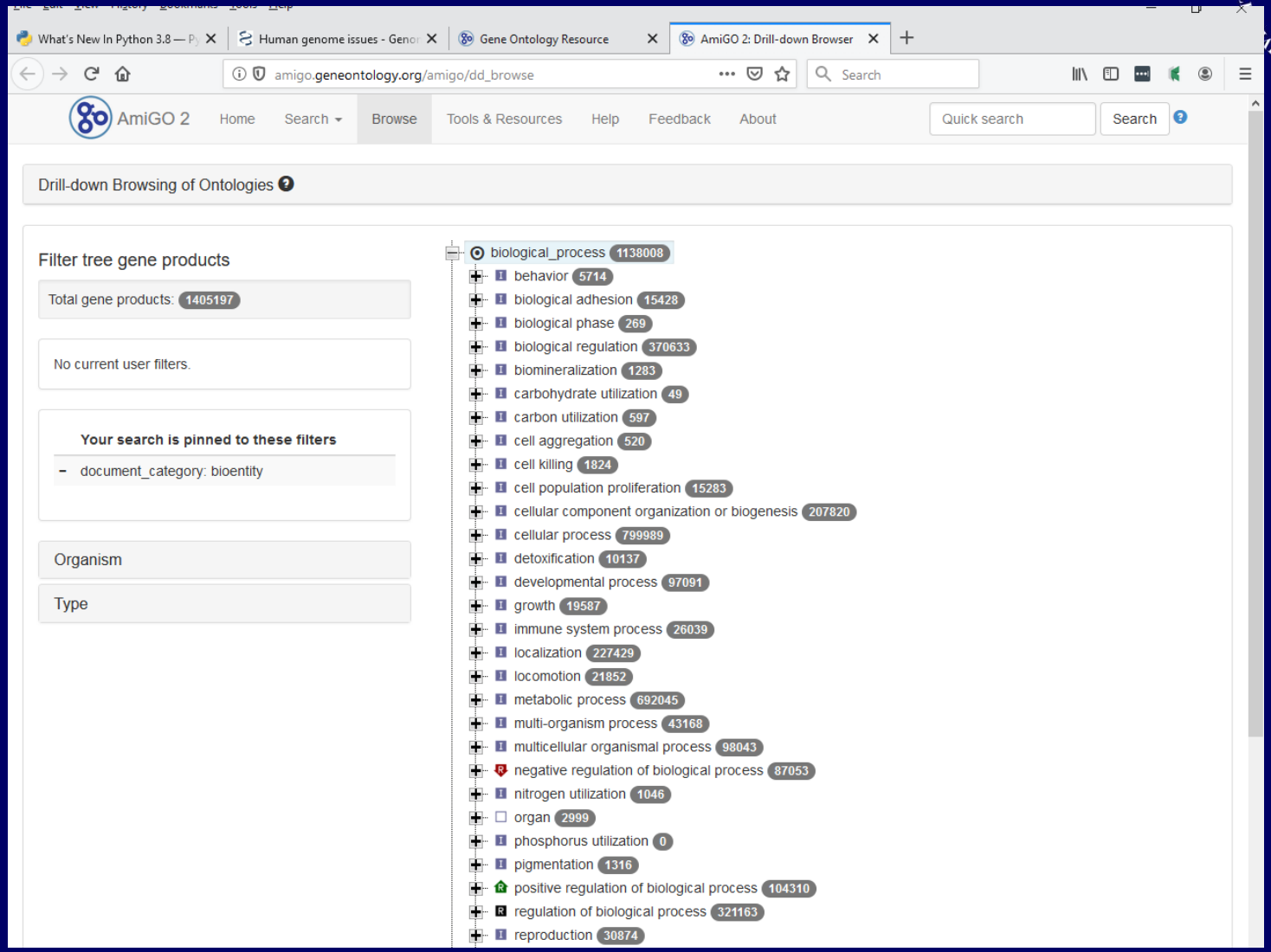

geneontology.org/stats.html
GENEONTOLOGY
Unifying Biology
About
Ontology
Annotations
Downloads
Help
ALLIANCE
Release statistics
In this section, the statistics of GO can be browsed for any of the releases archived in Zenodo from 2018-09-05 up to the current release 2020-10-09.
Statistics for release 2020-10
Ontology
Property | Value
Valid terms | 44264 (Δ = -8)
Obsoleted terms | 3017 (Δ = 46)
Merged terms | 2192 (Δ = 7)
Biological process terms | 28890
Molecular function terms | 11178
Cellular component terms | 4196
Annotations
Property | Value
Number of annotations | 8,049,377
Annotations for biological process | 3,060,065
Annotations for molecular function | 2,553,834
Annotations for cellular component | 2,435,478
Annotations for evidence PHYLO | 4,058,074
Annotations for evidence IEA | 2,038,699
Annotations for evidence EXP | 875,074
Annotations for evidence OTHER | 796,885
Annotations for evidence ND | 240,252
Annotations for evidence HTP | 40,393
Number of annotated scientific publications | 161,640
Gene products and species
Property | Value
Annotated gene products | 1,568,086
Annotated species | 4,666
Annotated species with over 1,000 annotations | 204
Changes in the GO resource over time

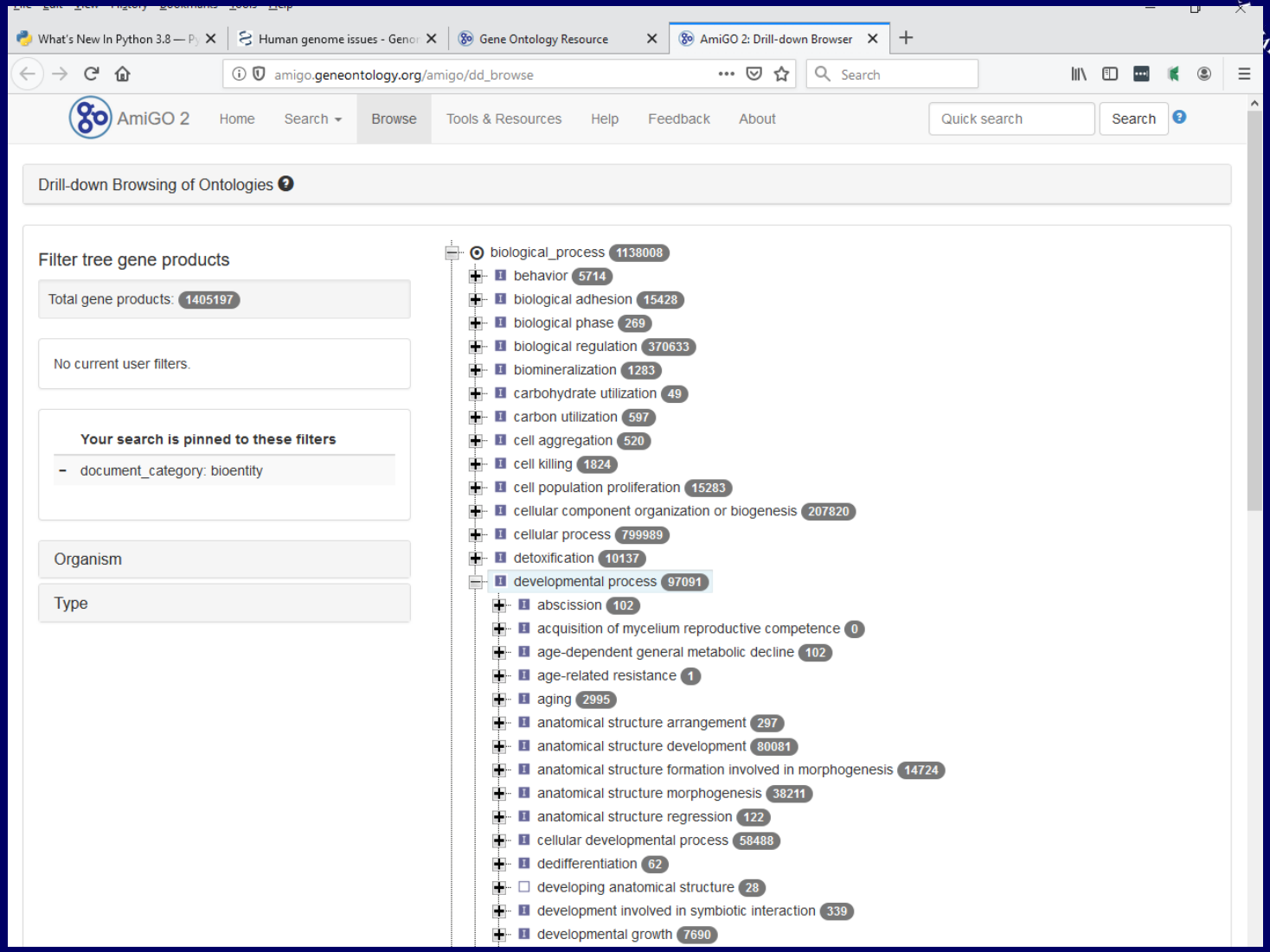

geneontology.org/stats.html
GENEONTOLOGY
Unifying Biology
About
Ontology
Annotations
Downloads
Help
ALLIANCE
hanges in the GO resource over time
Growth of annotated scientific publications over time
Number of GO terms by aspect
BP
MF
CC
Changes in GO terms between releases
created
merged
obsoleted
Number of annotations by evidence
Species filter:
All

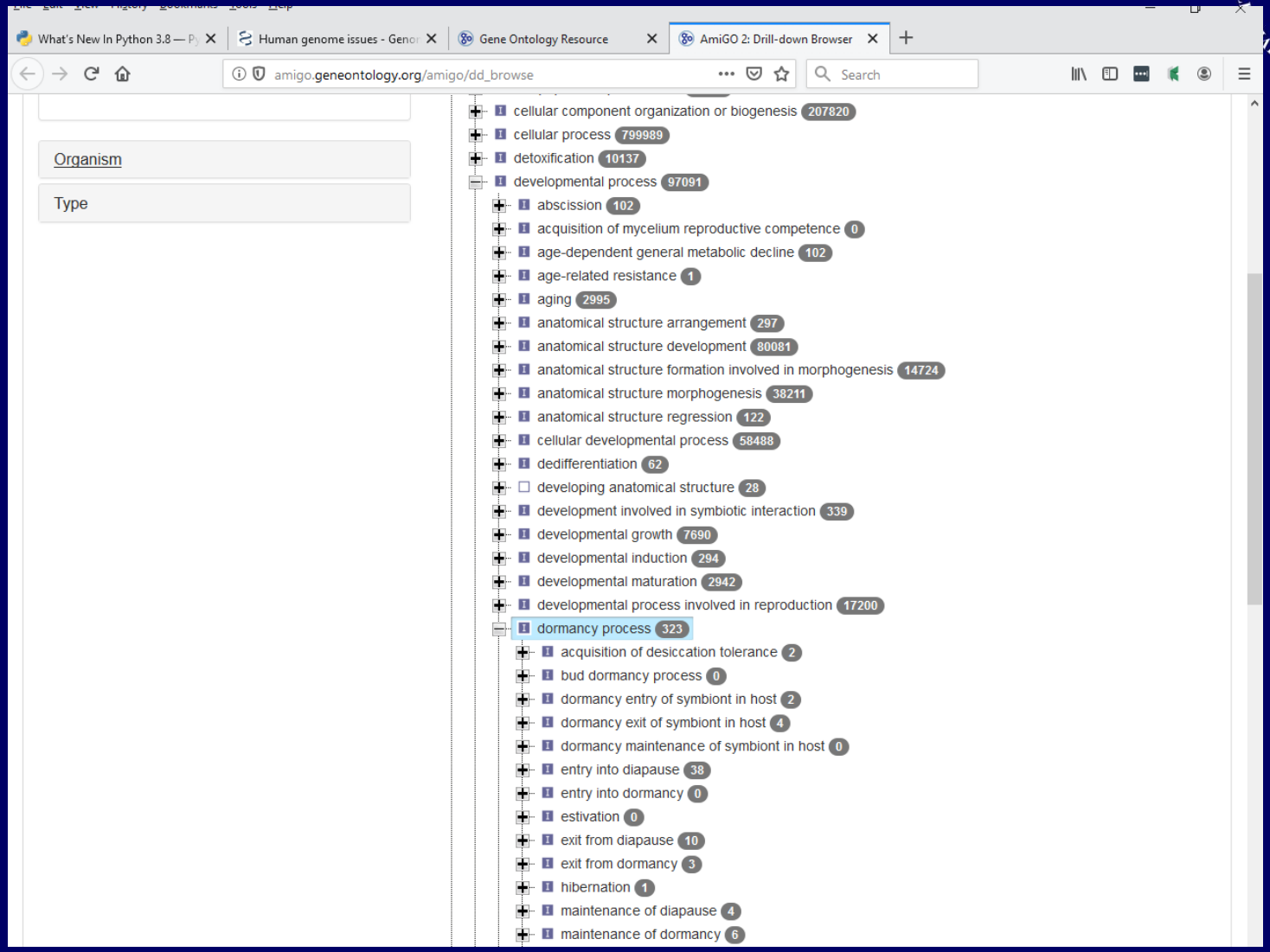

geneontology.org/stats.html
GENEONTOLOGY
Unifying Biology
About
Ontology
Annotations
Downloads
Help
ALLIANCE
BP
MF
CC
created
merged
obsoleted
Number of annotations by evidence
Species filter: All
10,000,000
8,000,000
6,000,000
4,000,000
2,000,000
0
2018-09
2018-11
2019-01
2019-03
2019-05
2019-07
2019-11
2020-01
2020-03
2020-05
2020-07
2020-09
PHYLO
IEA
ND
OTHER
EXP
HTP
Contact us
Citation and Terms of Use
Copyright © 1999-2020 Member of the
Open Biological Ontologies Foundry The Gene Ontology Consortium is
supported by a P41 grant from the National Human Genome Research
Institute (NHGRI) [grant U41 HG002273]

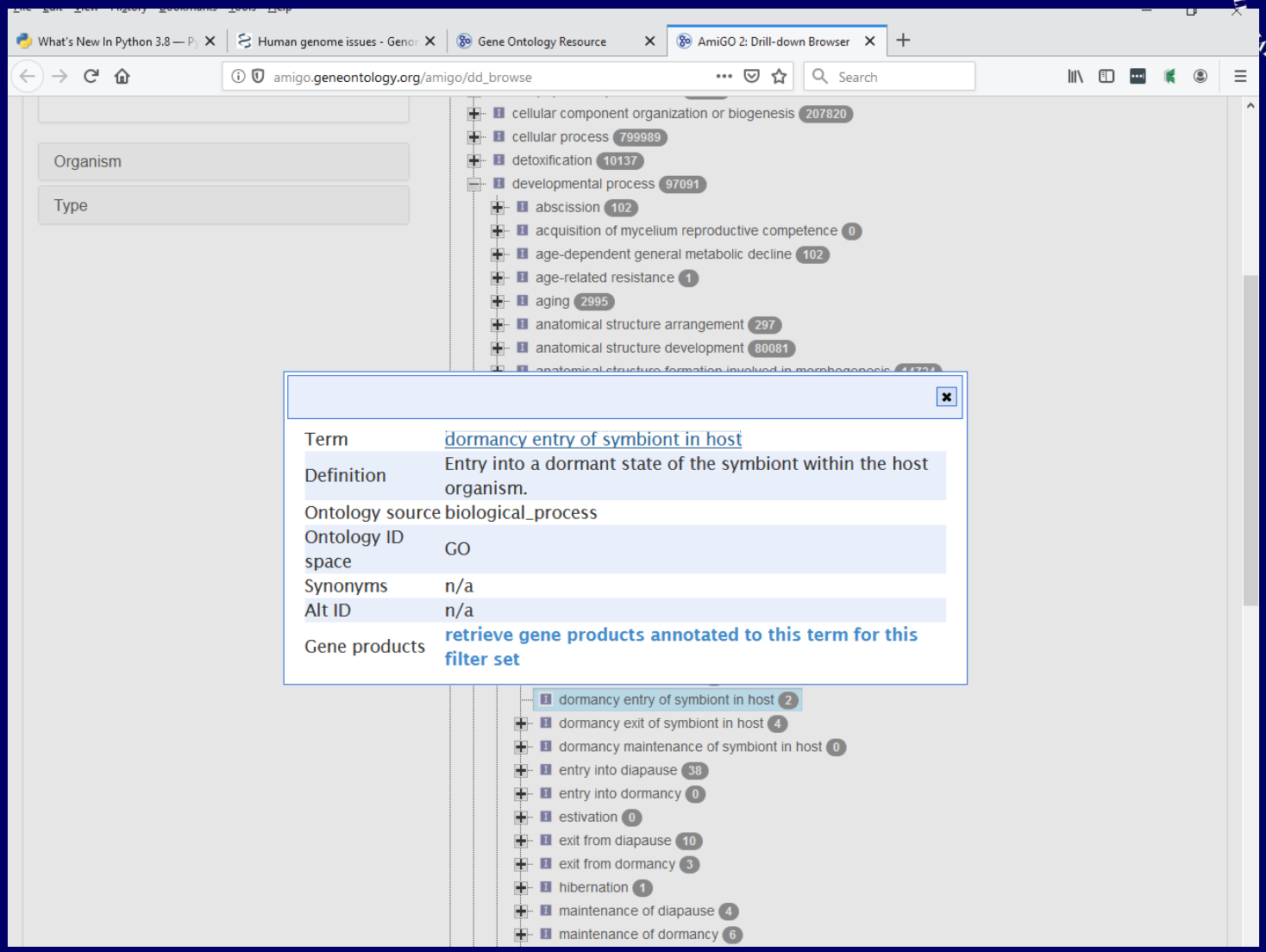

What's New In Python 3.8
Human genome issues - Geno
Gene Ontology Resource
geneontology.org
Search
GENEONTOLOGY
Unifying Biology
About
Ontology
Annotations
Downloads
Help
ALLIANCE
Search GO term or Gene Product in AmiGO ...
Homo sapiens
Examples
Launch
Any
Ontology
Gene Product
Hint: can use UniProt ID/AC, Gene Name, Gene Symbols, MOD IDs
ONTOLOGY
The network of biological classes describing the current best representation of the "universe" of biology. The molecular functions, cellular locations, and processes gene products may carry out.
GO Ontology Overview
Browse in AmiGO
Download
NEDD4
(evidence)
Ubiquitin-protein ligase activity
ANNOTATION
Statements, based on specific, traceable scientific evidence, asserting that a specific gene product is a real exemplar of a particular GO class.
GO Annotations Overview
View in AmiGO
Download
GO-CAM
GO Causal Activity Model (GO-CAM) provides a structured framework to link standard GO annotations into a more complete model of a biological system.
GO-CAM Overview
Browse GO-CAMs
Download
TOOL & GUIDE
Tools to curate, browse, search, visualize and download both the ontology and annotations. Bioinformatic Guides (Notebooks) and simple API access to integrate GO into your research.
GO Tools Overview
GO APIs Guide
GO GitHub
Home
Contact us
Citation and Terms of Use
Copyright © 1999-2019 Member of the Open Biological Ontologies Foundry
The Gene Ontology Consortium is supported by a P41 grant from the National Human Genome Research Institute (NHGRI) [grant U41 HG002273]

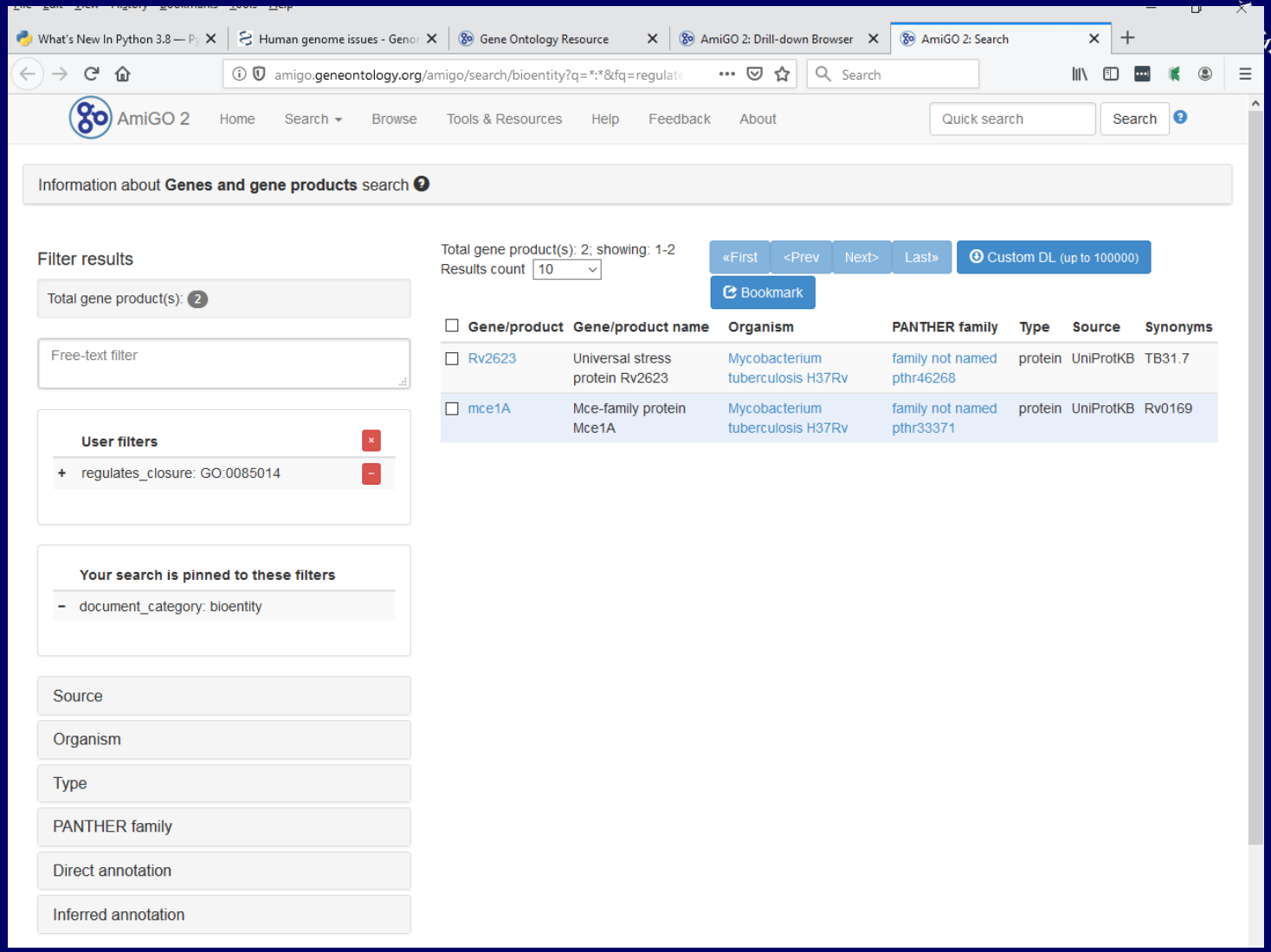

What's New In Python 3.8
Human genome issues - Geno
Gene Ontology Resource
AmiGO 2: Drill-down Browser
amigo.geneontology.org/amigo/dd_browse
Search
AmiGO 2
Home
Search
Browse
Tools & Resources
Help
Feedback
About
Quick search
Search
Drill-down Browsing of Ontologies
Filter tree gene products
Total gene products: 1405197
No current user filters.
Your search is pinned to these filters
document_category: bioentity
Organism
Type
biological_process 1138008
cellular_component 1073518
molecular_function 1068895
Cite this data • Terms of use • GO helpdesk
Copyright © 1999-2018 the Gene Ontology
DOI
Last file loaded on ???, see full details
AmiGO 2 version: 2.5.12 (amigo-production)

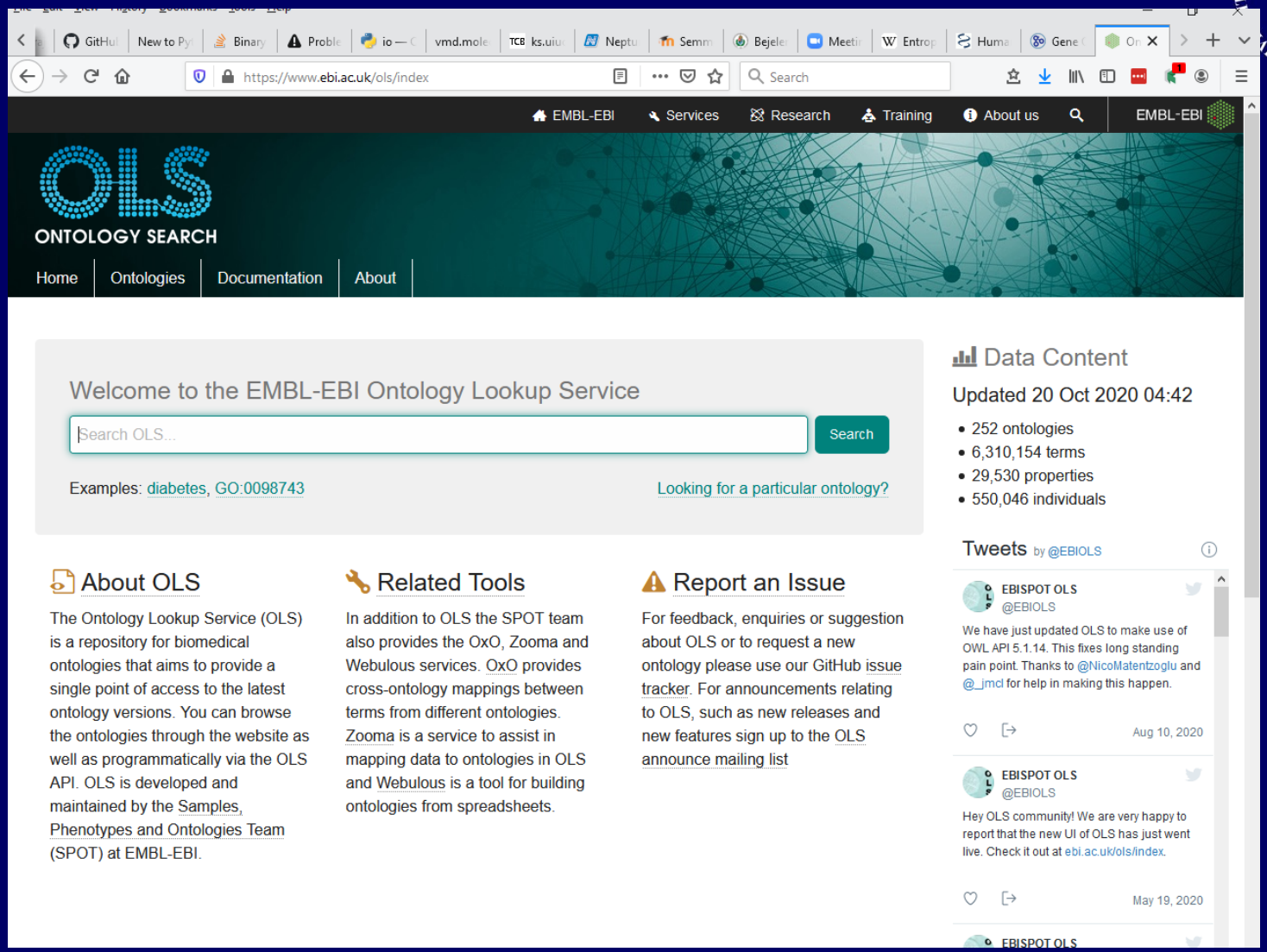

What's New In Python 3.8
Human genome issues - Geno
Gene Ontology Resource
AmiGO 2: Drill-down Browser
amigo.geneontology.org/amigo/dd_browse
Search
AmiGO 2
Home
Search
Browse
Tools & Resources
Help
Feedback
About
Quick search
Search
Drill-down Browsing of Ontologies
Filter tree gene products
Total gene products: 1405197
No current user filters.
Your search is pinned to these filters
document_category: bioentity
Organism
Type
biological_process 1138008
behavior 5714
biological adhesion 15428
biological phase 269
biological regulation 370633
biomineralization 1283
carbohydrate utilization 49
carbon utilization 597
cell aggregation 520
cell killing 1824
cell population proliferation 15283
cellular component organization or biogenesis 207820
cellular process 799989
detoxification 10137
developmental process 97091
growth 19587
immune system process 26039
localization 227429
locomotion 21852
metabolic process 692045
multi-organism process 43168
multicellular organismal process 98043
negative regulation of biological process 87053
nitrogen utilization 1046
organ 2999
phosphorus utilization 0
pigmentation 1316
positive regulation of biological process 104310
regulation of biological process 321163
reproduction 30874

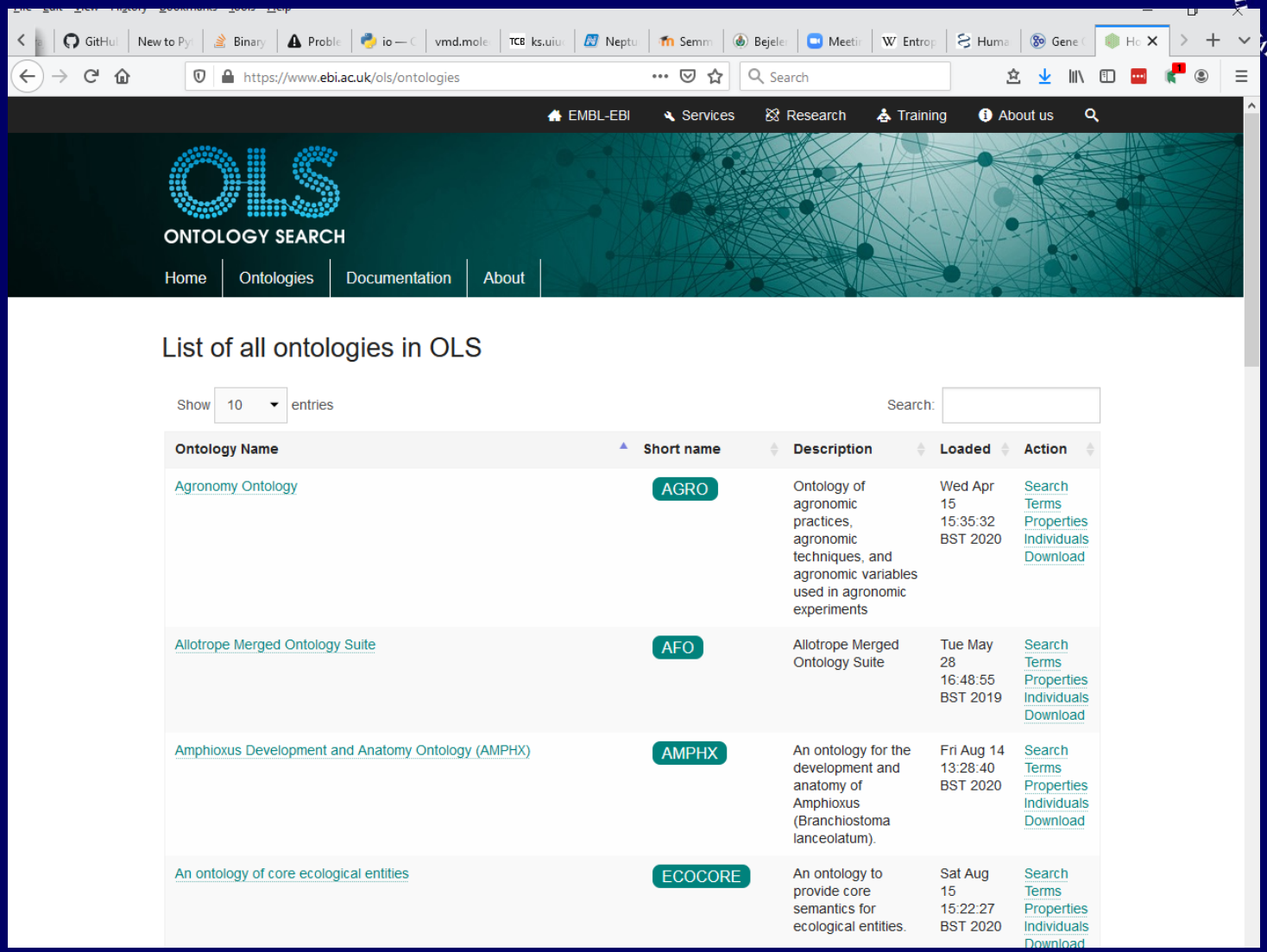

What's New In Python 3.8
Human genome issues - Geno
Gene Ontology Resource
AmiGO 2: Drill-down Browser
amigo.geneontology.org/amigo/dd_browse
Search
AmiGO 2
Home
Search
Browse
Tools & Resources
Help
Feedback
About
Quick search
Search
Drill-down Browsing of Ontologies
Filter tree gene products
Total gene products: 1405197
No current user filters.
Your search is pinned to these filters
document_category: bioentity
Organism
Type
biological_process 1138008
behavior 5714
biological adhesion 15428
biological phase 269
biological regulation 370633
biomineralization 1283
carbohydrate utilization 49
carbon utilization 597
cell aggregation 520
cell killing 1824
cell population proliferation 15283
cellular component organization or biogenesis 207820
cellular process 799989
detoxification 10137
developmental process 97091
abscission 102
acquisition of mycelium reproductive competence 0
age-dependent general metabolic decline 102
age-related resistance 1
aging 2995
anatomical structure arrangement 297
anatomical structure development 80081
anatomical structure formation involved in morphogenesis 14724
anatomical structure morphogenesis 38211
anatomical structure regression 122
cellular developmental process 58488
dedifferentiation 62
developing anatomical structure 28
development involved in symbiotic interaction 339
developmental growth 7690


What's New In Python 3.8
Human genome issues - Geno
Gene Ontology Resource
AmiGO 2: Drill-down Browser
amigo.geneontology.org/amigo/dd_browse
Search
Organism
Type
cellular component organization or biogenesis 207820
cellular process 799989
detoxification 10137
developmental process 97091
abscission 102
acquisition of mycelium reproductive competence 0
age-dependent general metabolic decline 102
age-related resistance 1
aging 2995
anatomical structure arrangement 297
anatomical structure development 80081
anatomical structure formation involved in morphogenesis 14724
anatomical structure morphogenesis 38211
anatomical structure regression 122
cellular developmental process 58488
dedifferentiation 62
developing anatomical structure 28
development involved in symbiotic interaction 339
developmental growth 7690
developmental induction 294
developmental maturation 2942
developmental process involved in reproduction 17200
dormancy process 323
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
entry into diapause 38
entry into dormancy 0
estivation 0
exit from diapause 10
exit from dormancy 3
hibernation 1
maintenance of diapause 4
maintenance of dormancy 6


What's New In Python 3.8
Human genome issues - Geno
Gene Ontology Resource
AmiGO 2: Drill-down Browser
amigo.geneontology.org/amigo/dd_browse
Search
Organism
Type
cellular component organization or biogenesis 207820
cellular process 799989
detoxification 10137
developmental process 97091
abscission 102
acquisition of mycelium reproductive competence 0
age-dependent general metabolic decline 102
age-related resistance 1
aging 2995
anatomical structure arrangement 297
anatomical structure development 80081
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
dormancy entry of symbiont in host 2
dormancy exit of symbiont in host 4
dormancy maintenance of symbiont in host 0
entry into diapause 38
entry into dormancy 0
estivation 0
exit from diapause 10
exit from dormancy 3
hibernation 1
maintenance of diapause 4
maintenance of dormancy 6


What's New In Python 3.8
Human genome issues - Geno
Gene Ontology Resource
AmiGO 2: Drill-down Browser
AmiGO 2: Search
amigo.geneontology.org/amigo/search/bioentity?q=*:*&fq=regulate
Search
AmiGO 2
Home
Search
Browse
Tools & Resources
Help
Feedback
About
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
family not named pthr46268
protein
UniProtKB
TB31.7
mce1A
Mce-family protein Mce1A
Mycobacterium tuberculosis H37Rv
family not named pthr33371
protein
UniProtKB
Rv0169


https://www.ebi.ac.uk/ols/index
Search
EMBL-EBI
Services
Research
Training
About us
EMBL-EBI
OLS
ONTOLOGY SEARCH
Home
Ontologies
Documentation
About
Welcome to the EMBL-EBI Ontology Lookup Service
Search OLS...
Search
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 the OxO, Zooma and Webulous services. OxO provides cross-ontology mappings between terms from different ontologies. Zooma is a service to assist in mapping data to ontologies in OLS and Webulous is a tool for building ontologies from spreadsheets.
Report an Issue
For feedback, enquiries or suggestion about OLS or to request a new ontology please use our GitHub issue tracker. For announcements relating to OLS, such as new releases and new features sign up to the OLS announce mailing list
Data Content
Updated 20 Oct 2020 04:42
252 ontologies
6,310,154 terms
29,530 properties
550,046 individuals
Tweets by @EBIOLS
EBISPOT OLS
@EBIOLS
We have just updated OLS to make use of OWL API 5.1.14. This fixes long standing pain point. Thanks to @NicoMatentzoglu and @_jmcl for help in making this happen.
Aug 10, 2020
EBISPOT OLS
@EBIOLS
Hey OLS community! We are very happy to report that the new UI of OLS has just went live. Check it out at ebi.ac.uk/ols/index.
May 19, 2020


https://www.ebi.ac.uk/ols/ontologies
EMBL-EBI
Services
Research
Training
About us
OLS
ONTOLOGY SEARCH
Home
Ontologies
Documentation
About
List of all ontologies in OLS
Show 10 entries
Search:
Ontology Name
Short name
Description
Loaded
Action
Agronomy Ontology
AGRO
Ontology of agronomic practices, agronomic techniques, and agronomic variables used in agronomic experiments
Wed Apr 15 15:35:32 BST 2020
Search Terms Properties Individuals Download
Allotrope Merged Ontology Suite
AFO
Allotrope Merged Ontology Suite
Tue May 28 16:48:55 BST 2019
Search Terms Properties Individuals Download
Amphioxus Development and Anatomy Ontology (AMPHX)
AMPHX
An ontology for the development and anatomy of Amphioxus (Branchiostoma lanceolatum).
Fri Aug 14 13:28:40 BST 2020
Search Terms Properties Individuals Download
An ontology of core ecological entities
ECOCORE
An ontology to provide core semantics for ecological entities.
Sat Aug 15 15:22:27 BST 2020
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
  - Minden feladatra külön eszköz
- Hierarchikus – RuleManager
  - Ez osztja ki a feladatokat és ellenőrzi az eredményt
- A számítások „computer farm”-on mennek
  - ~1000 node, közös adatbázist ér el

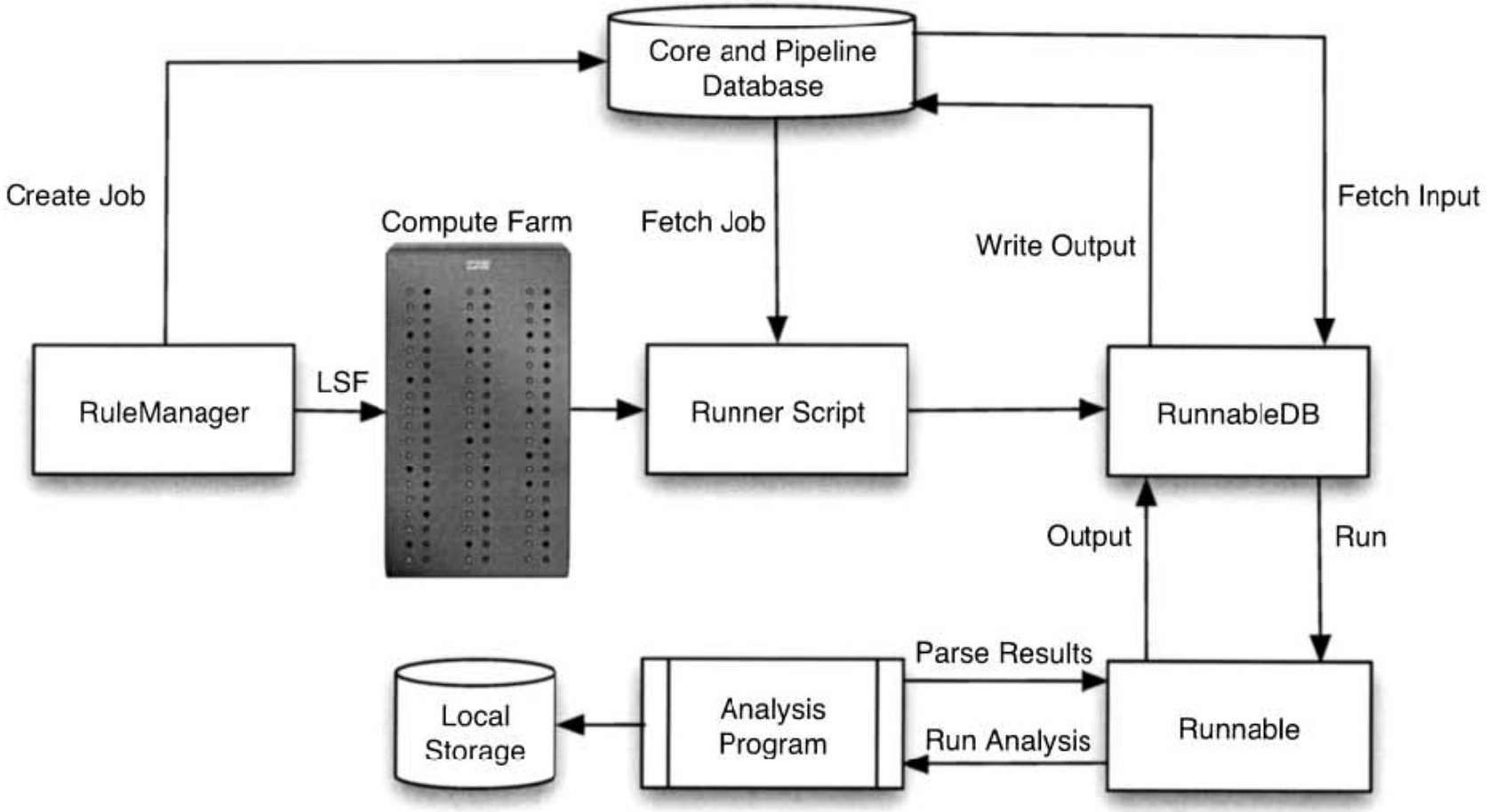
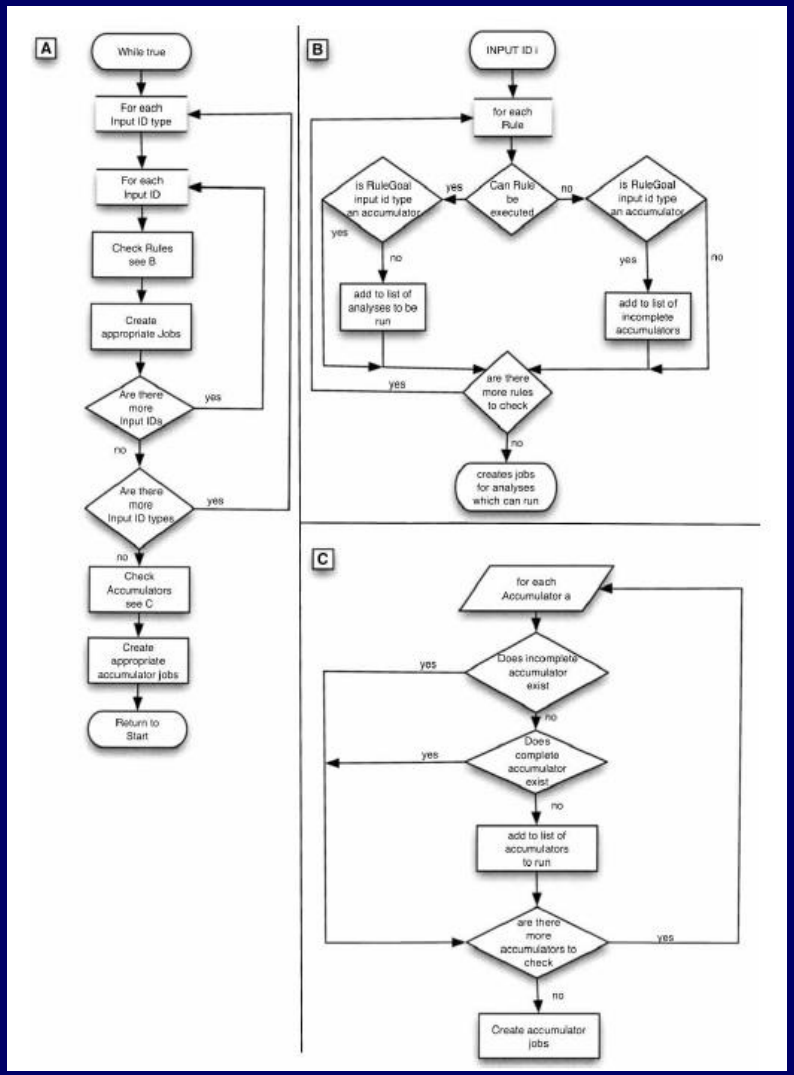


Figure 1 Ensembl pipeline system overview: The RuleManager uses LSF to submit analysis jobs to the compute farm. When an individual job starts executing on a remote node, the Runner script fetches the job information from the database and recreates the Job object. This in turn creates a RunnableDB and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



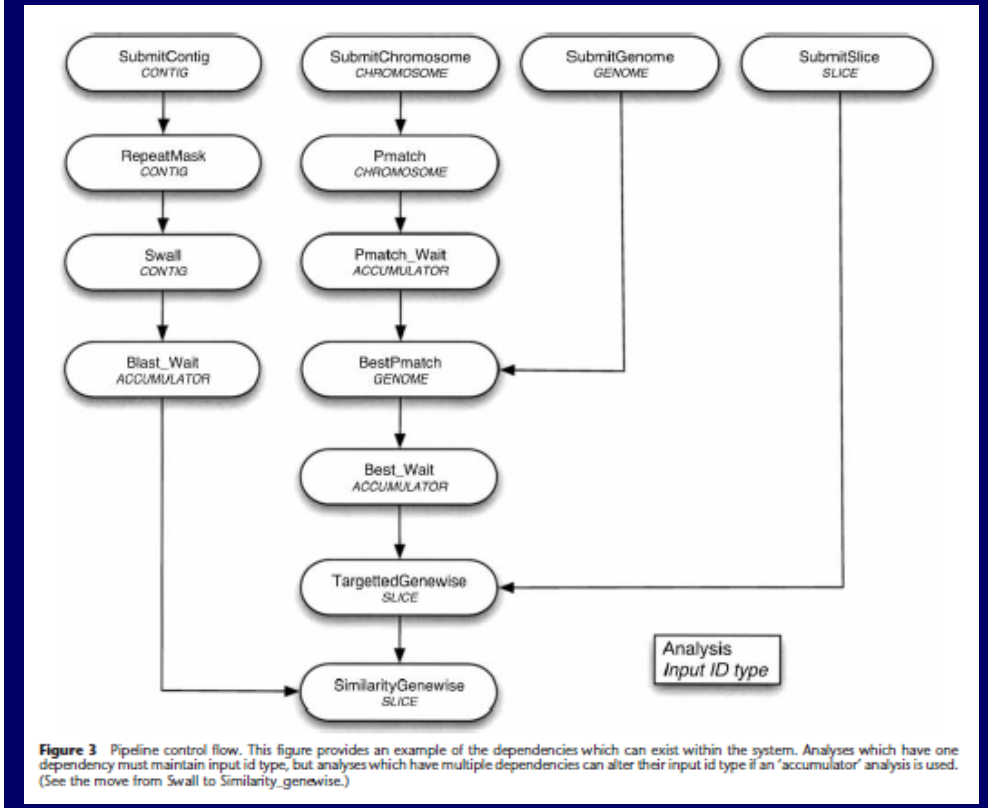


Figure 3 Pipeline control flow. This figure provides an example of the dependencies which can exist within the system. Analyses which have one dependency must maintain input id type, but analyses which have multiple dependencies can alter their input id type if an 'accumulator' analysis is used. (See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

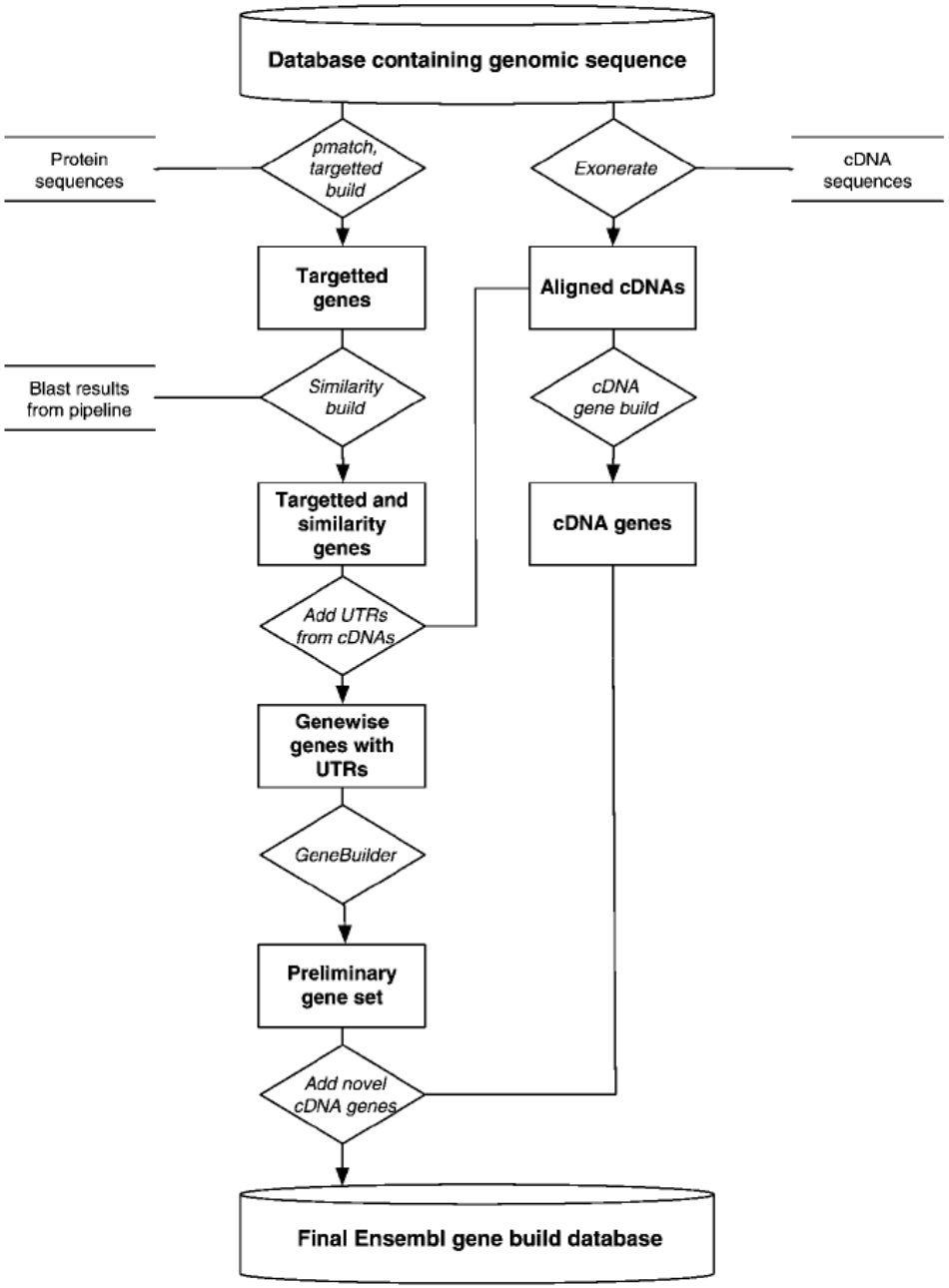
- A pmatch jelentősen lecsökkenti a vizsgálandó szegmensek hosszát
- Utána genewise – fehérje szintű illesztés splice-helyekkel és frame eltolással
- Viszont elég lassú
- *miniseq*: a megtalált nyers exonokat megtoldjuk 200 bázissal mindkét irányba
- Ez kezelhető méretűvé csökkenti a szekvenci hosszát

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 megszűrjük és csak a legjobb 10-et tartjuk meg (ritka eset)
4. Ismét klaszterezzük a géneket, ha az előző lépés új klasztereket hozott volna létre


Database containing genomic sequence
Protein sequences
pmatch, targetted build
Exonerate
cDNA sequences
Targetted genes
Aligned cDNAs
Blast results from pipeline
Similarity build
cDNA gene build
Targetted and similarity genes
cDNA genes
Add UTRs from cDNAs
Genewise genes with UTRs
GeneBuilder
Preliminary gene set
Add novel cDNA genes
Final Ensembl gene build database

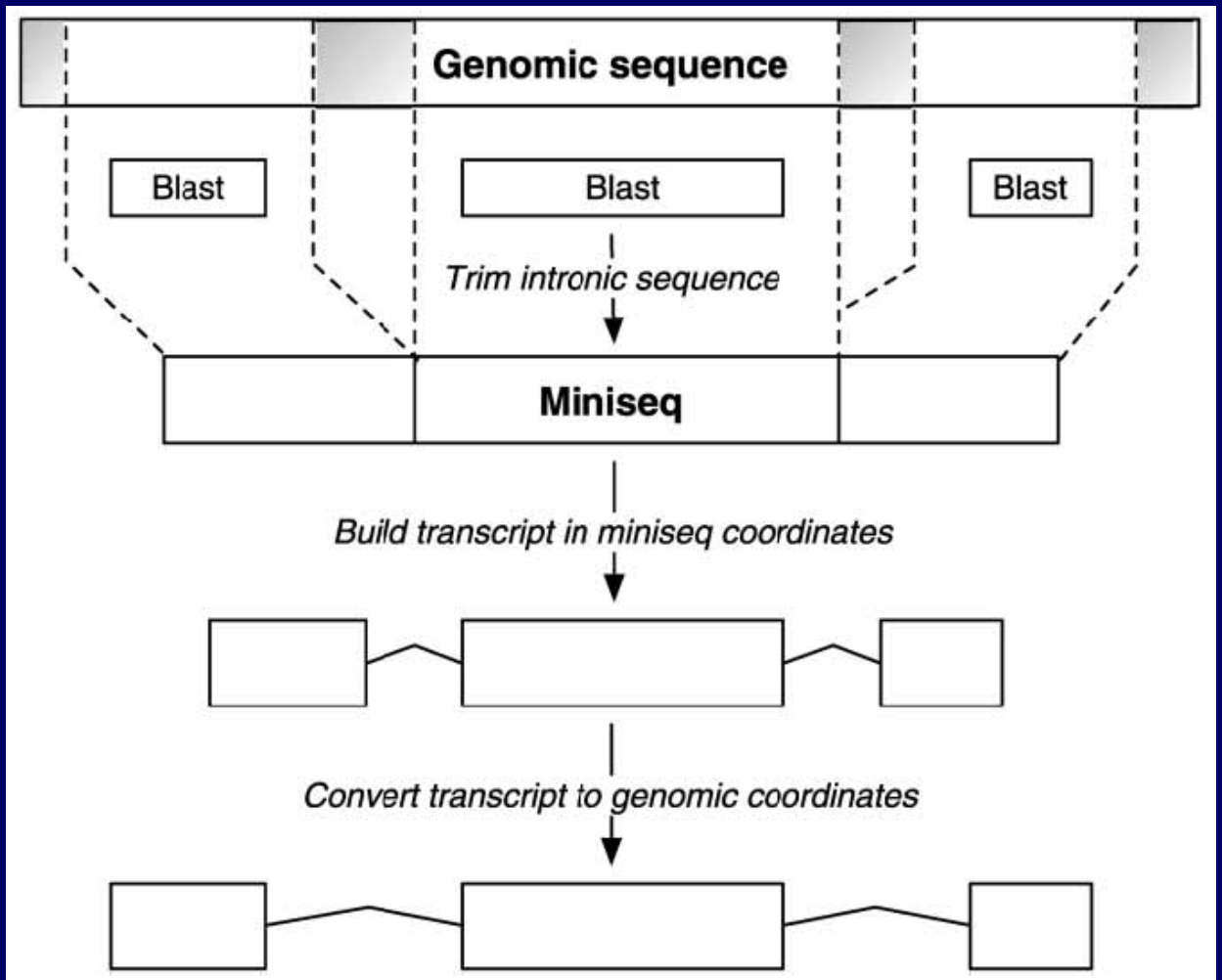
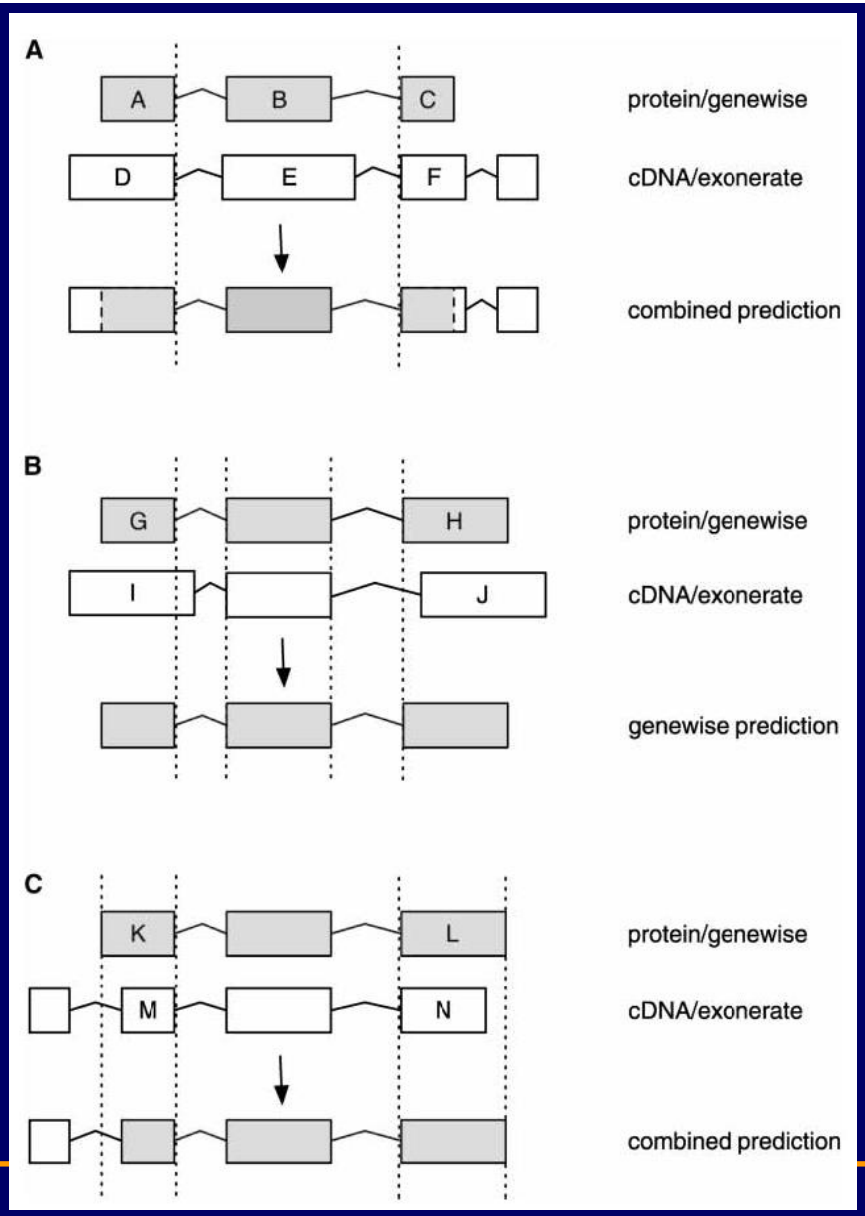
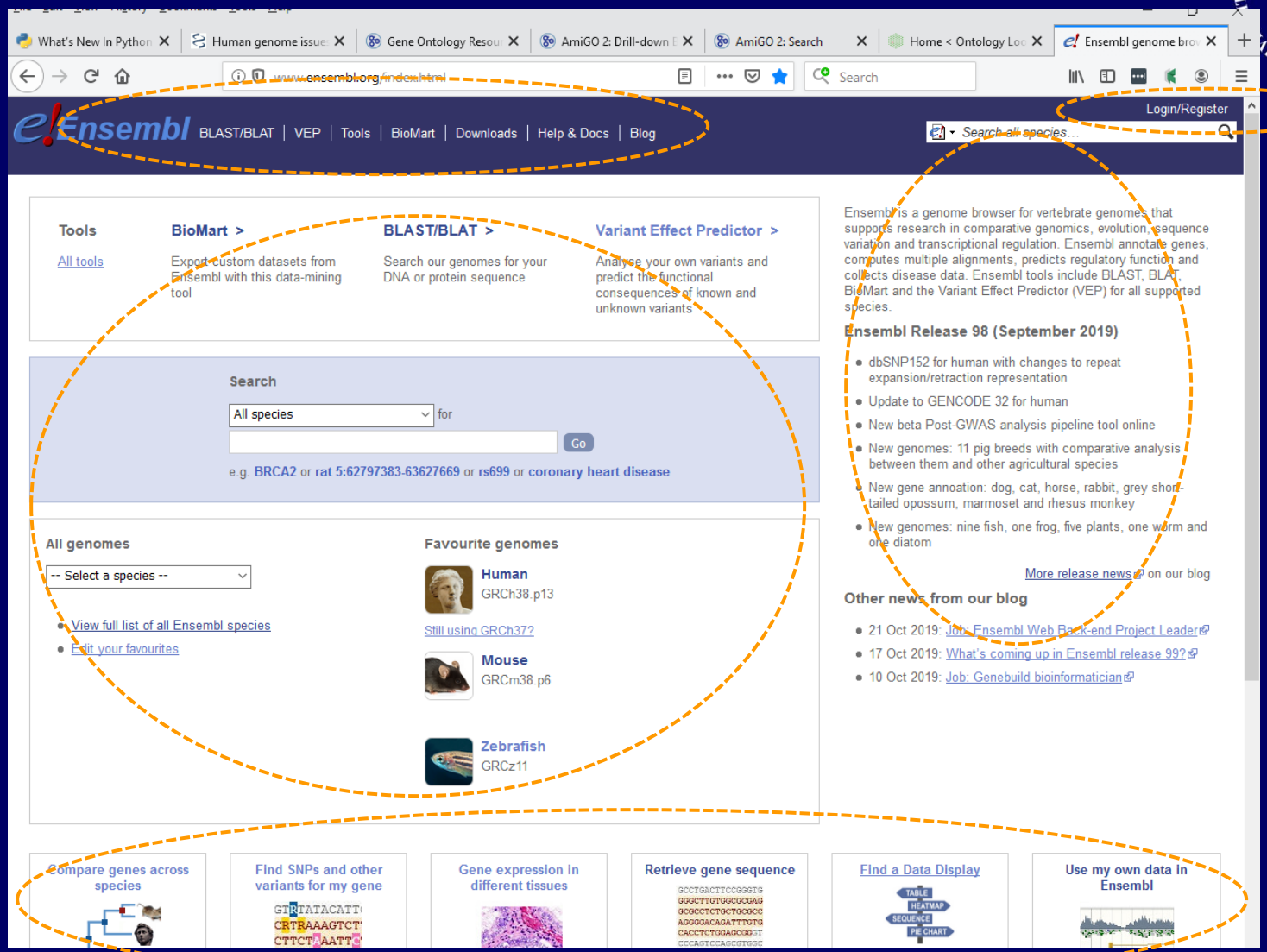


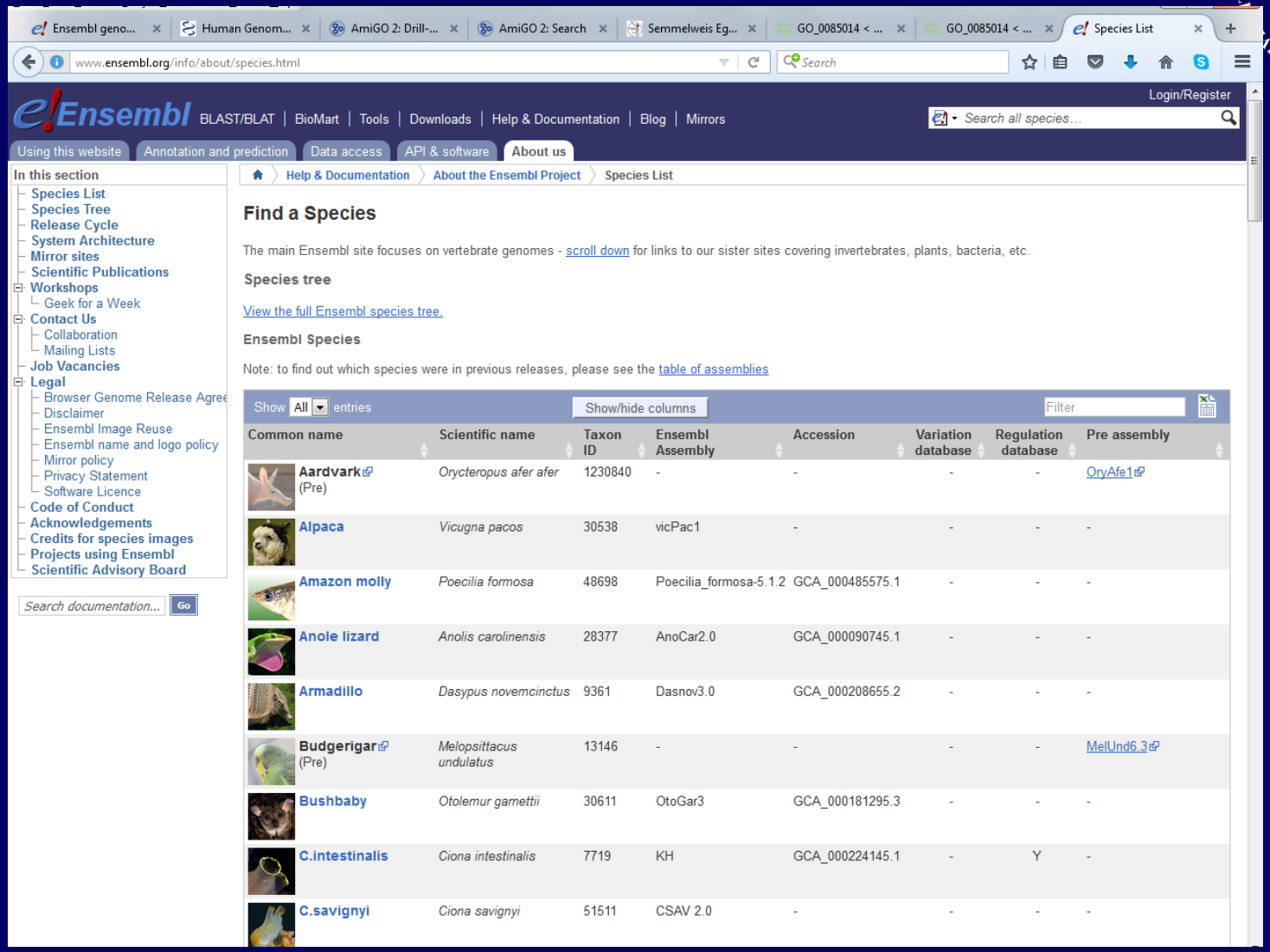
Figure 2 The Miniseq: We use a miniseq representation of genomic sequence in various stages of the gene build in order to reduce search space and increase processing speed. We BLAST a sequence of interest against a genomic region and pad the resulting hits with 200 bp. We then join the padded hits together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

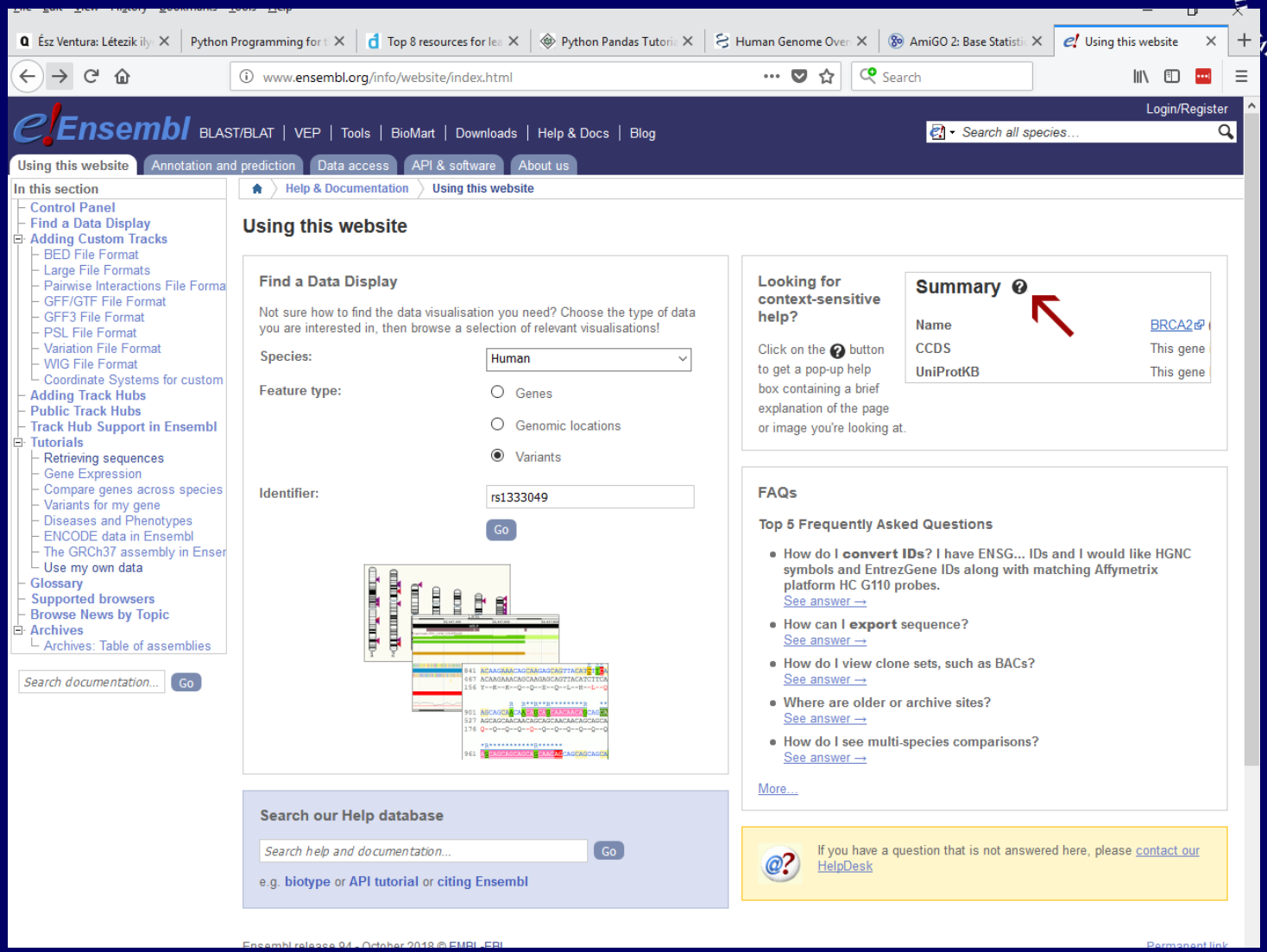

A
A
B
C
protein/genewise
D
E
F
cDNA/exonerate
combined prediction
B
G
H
protein/genewise
I
J
cDNA/exonerate
genewise prediction
C
K
L
protein/genewise
M
N
cDNA/exonerate
combined prediction

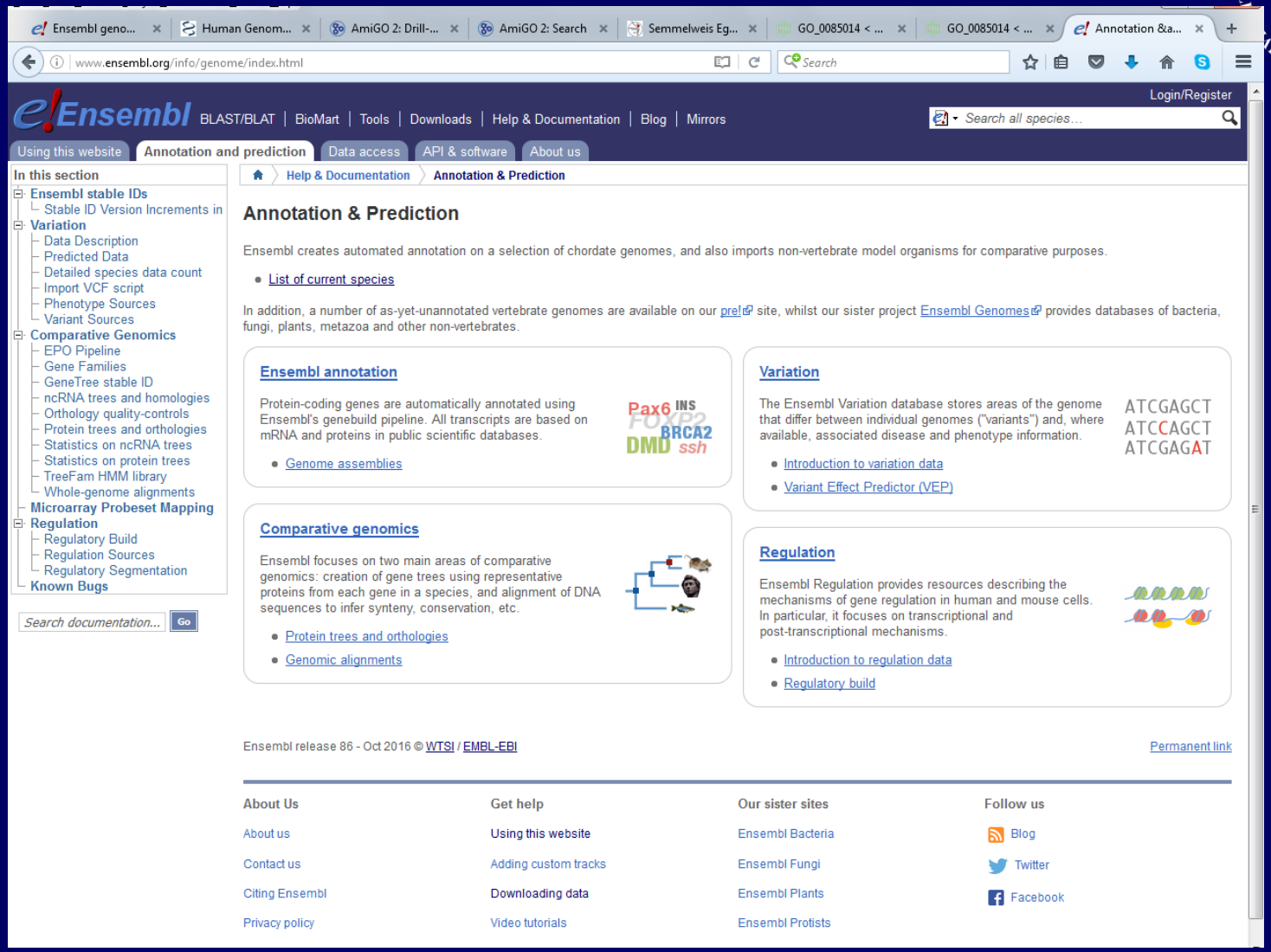
Az ENSEMBL felület

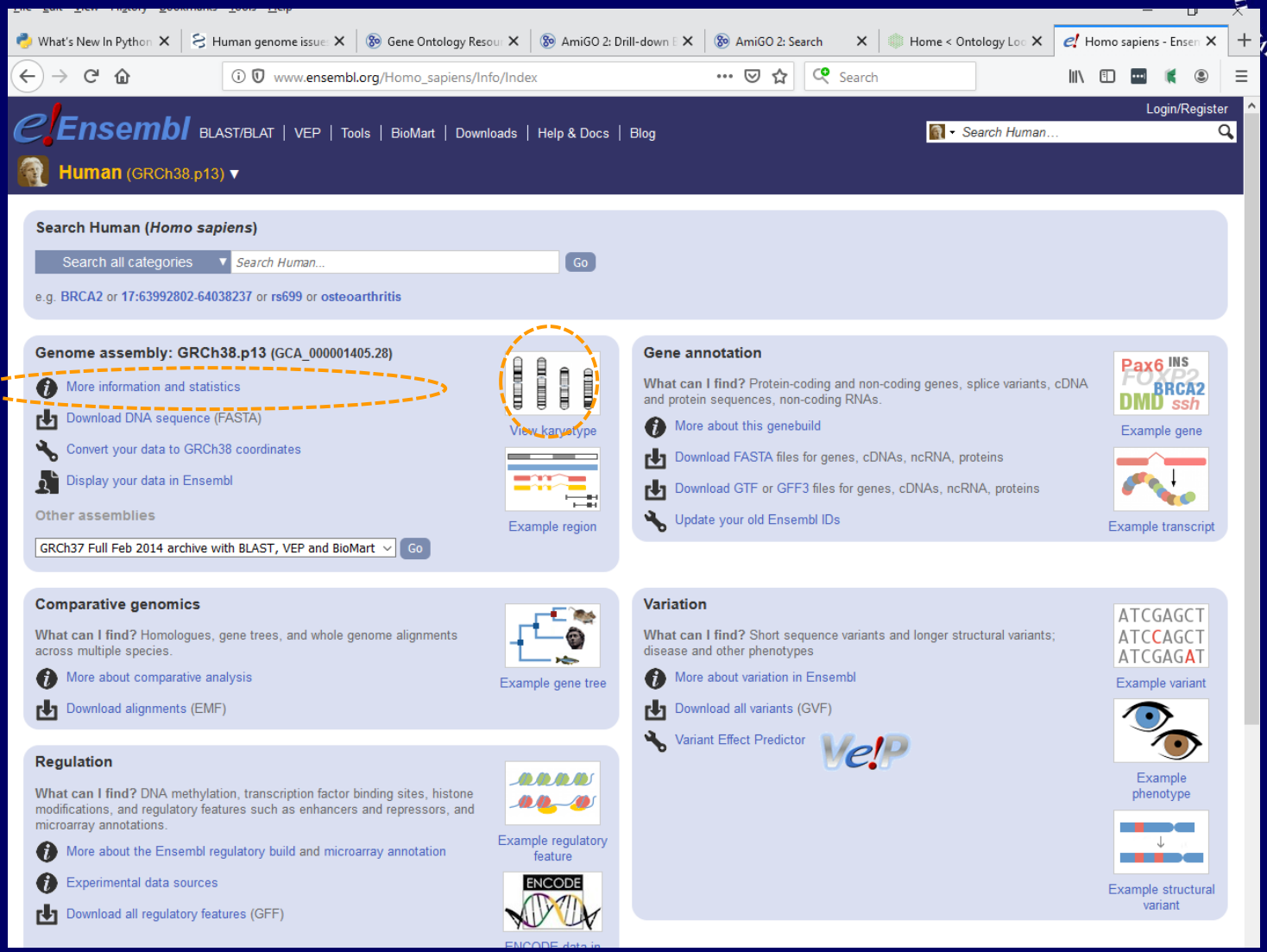
- Honlap: http://www.ensembl.org/index.html
- Többszáz élőlény – többségében gerinces
- Ingyenesen elérhető, folyamatosan fejlesztik
- Közös Európai fenntartású kezdeményezés
- Több módon is elérhetők az adatok:
  - Web, BioMart, FTP
- Adatok feltöltése is lehetséges

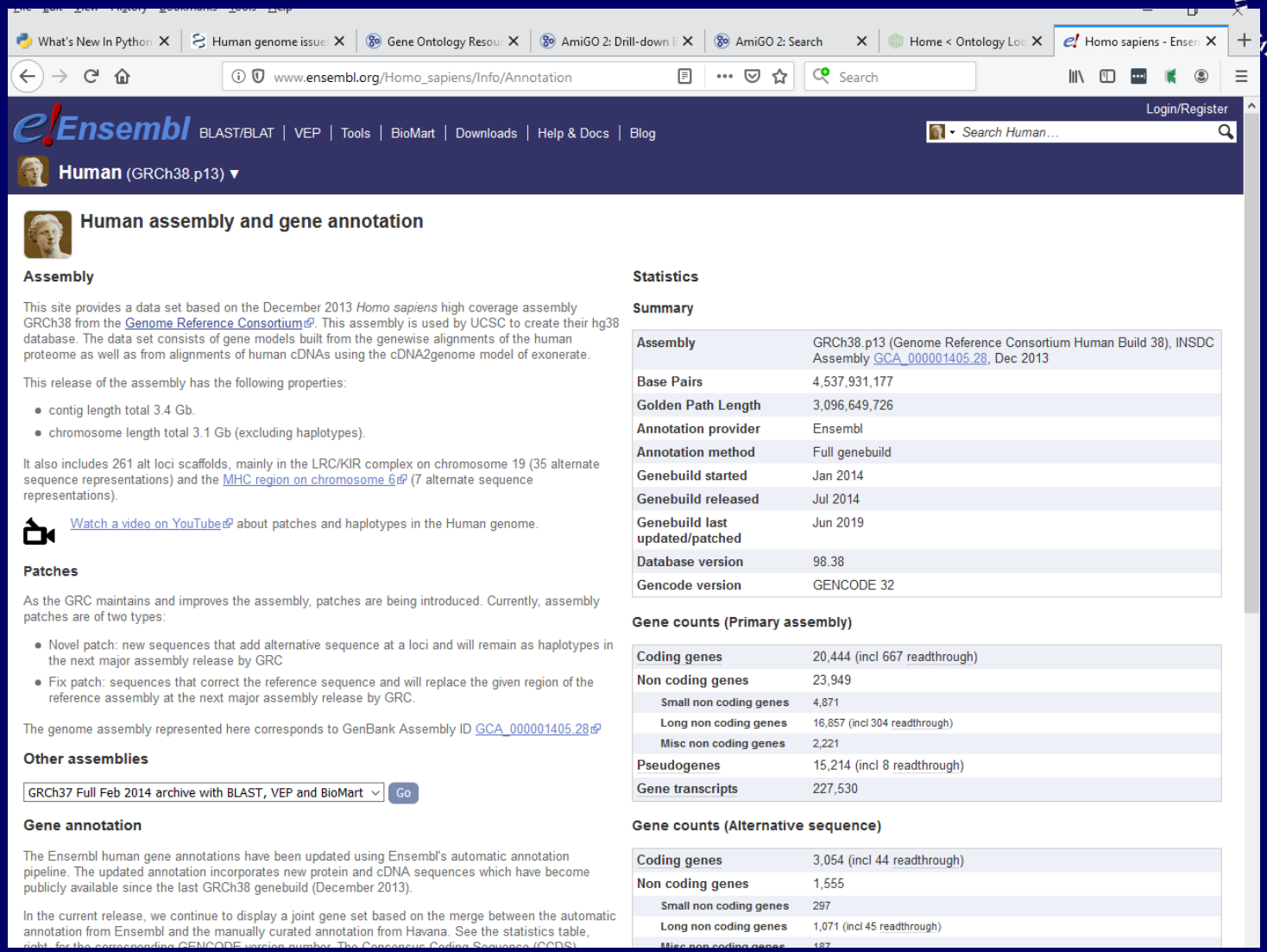

Ensembl
BLAST/BLAT | VEP | Tools | BioMart | Downloads | Help & Docs | Blog
Login/Register
Search all species...
Tools
All tools
BioMart >
Export custom datasets from Ensembl with this data-mining tool
BLAST/BLAT >
Search our genomes for your DNA or protein sequence
Variant Effect Predictor >
Analyse your own variants and predict the functional consequences of known and unknown variants
Ensembl is a genome browser for vertebrate genomes that supports research in comparative genomics, evolution, sequence variation and transcriptional regulation. Ensembl annotate genes, computes multiple alignments, predicts regulatory function and collects disease data. Ensembl tools include BLAST, BLAT, BioMart and the Variant Effect Predictor (VEP) for all supported species.
Search
All species
for
Go
e.g. BRCA2 or rat 5:62797383-63627669 or rs699 or coronary heart disease
Ensembl Release 98 (September 2019)
dbSNP152 for human with changes to repeat expansion/retraction representation
Update to GENCODE 32 for human
New beta Post-GWAS analysis pipeline tool online
New genomes: 11 pig breeds with comparative analysis between them and other agricultural species
New gene annoation: dog, cat, horse, rabbit, grey short-tailed opossum, marmoset and rhesus monkey
New genomes: nine fish, one frog, five plants, one worm and one diatom
More release news on our blog
All genomes
-- Select a species --
View full list of all Ensembl species
Edit your favourites
Favourite genomes
Human
GRCh38.p13
Still using GRCh37?
Mouse
GRCm38.p6
Zebrafish
GRCz11
Other news from our blog
21 Oct 2019: Job: Ensembl Web Back-end Project Leader
17 Oct 2019: What's coming up in Ensembl release 99?
10 Oct 2019: Job: Genebuild bioinformatician
Compare genes across species
Find SNPs and other variants for my gene
Gene expression in different tissues
Retrieve gene sequence
Find a Data Display
Use my own data in Ensembl

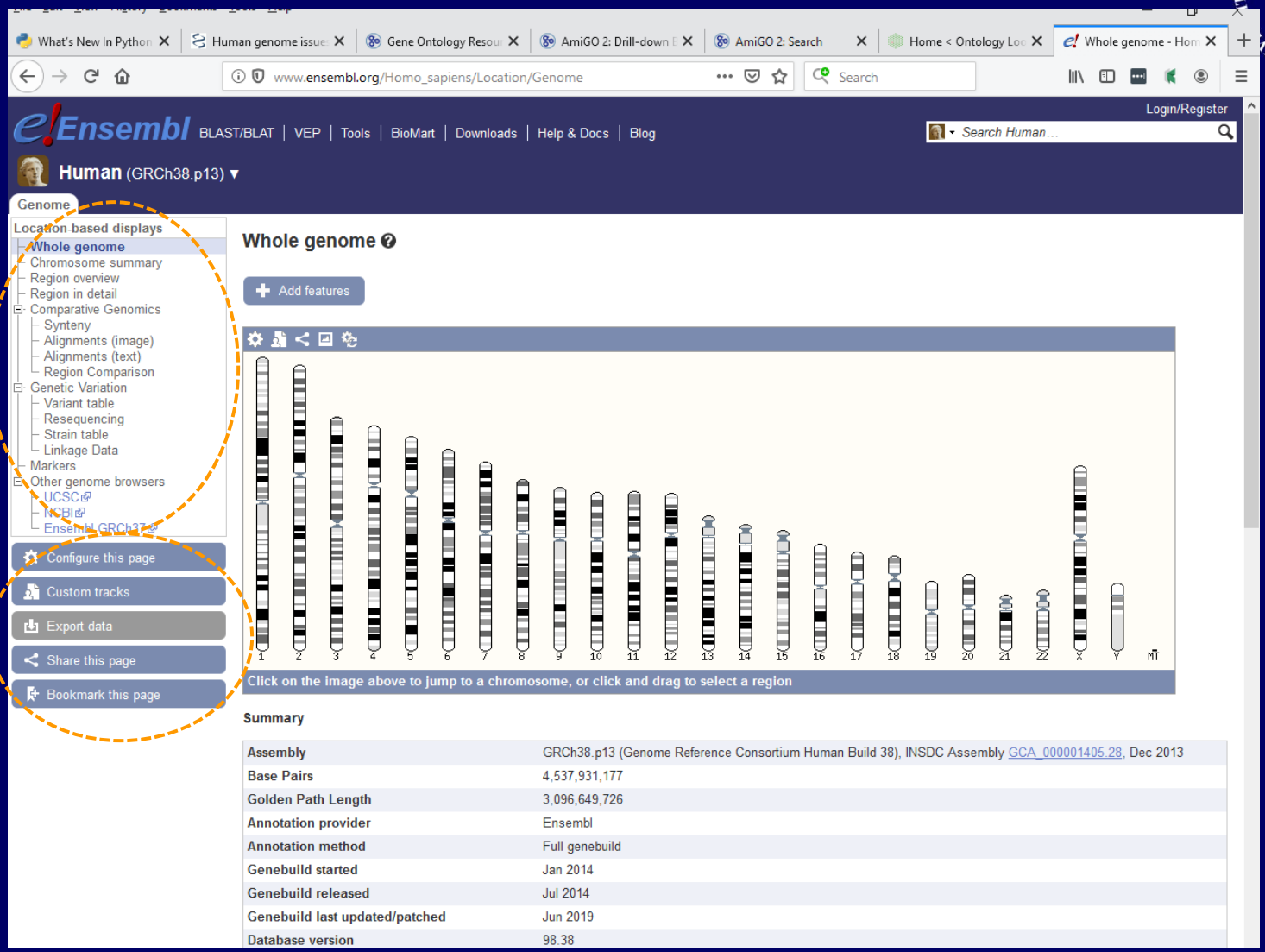

Ensembl geno...
Human Genom...
AmiGO 2: Drill-...
AmiGO 2: Search
Semmelweis Eg...
GO_0085014 < ...
GO_0085014 < ...
Species List
www.ensembl.org/info/about/species.html
Search
Ensembl
BLAST/BLAT | BioMart | Tools | Downloads | Help & Documentation | Blog | Mirrors
Login/Register
Search all species...
Using this website
Annotation and prediction
Data access
API & software
About us
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
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
Show All entries
Show/hide columns
Filter
Common name
Scientific name
Taxon ID
Ensembl Assembly
Accession
Variation database
Regulation database
Pre assembly
Aardvark (Pre)
Oryctéropus afer afer
1230840
-
-
-
-
OryAfe1
Alpaca
Vicugna pacos
30538
vicPac1
-
-
-
-
Amazon molly
Poecilia formosa
48698
Poecilia_formosa-5.1.2
GCA_000485575.1
-
-
-
Anole lizard
Anolis carolinensis
28377
AnoCar2.0
GCA_000090745.1
-
-
-
Armadillo
Dasypus novemcinctus
9361
Dasnov3.0
GCA_000208655.2
-
-
-
Budgerigar (Pre)
Melopsittacus undulatus
13146
-
-
-
-
MelUnd6.3
Bushbaby
Otolemur garnettii
30611
OtoGar3
GCA_000181295.3
-
-
-
C.intestinalis
Ciona intestinalis
7719
KH
GCA_000224145.1
-
Y
-
C.savignyi
Ciona savignyi
51511
CSAV 2.0
-
-
-
-

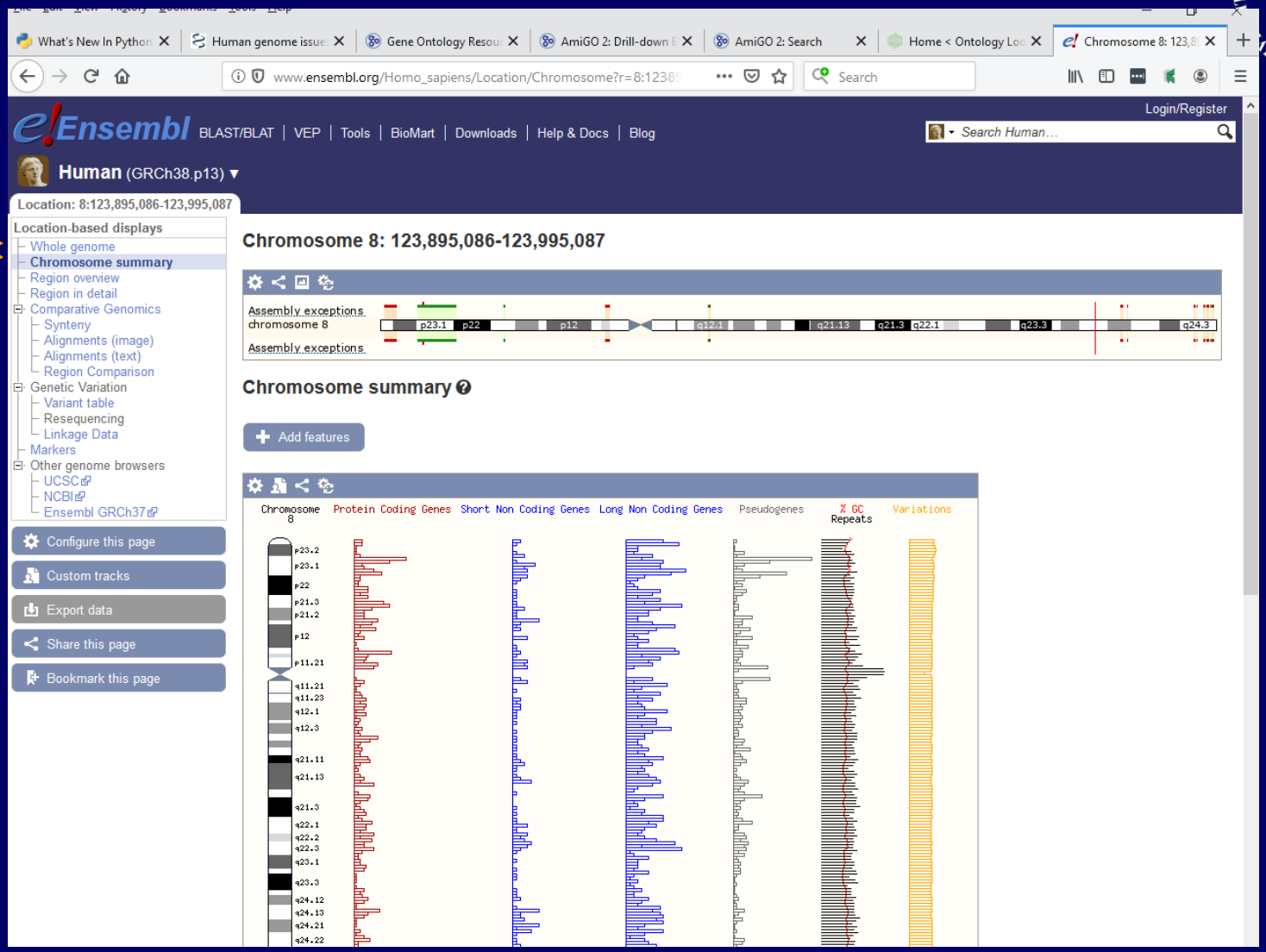

Using this website
www.ensembl.org/info/website/index.html
Ensembl
BLAST/BLAT | VEP | Tools | BioMart | Downloads | Help & Docs | Blog
Login/Register
Search all species...
Using this website
Annotation and prediction
Data access
API & software
About us
In this section
Control Panel
Find a Data Display
Adding Custom Tracks
BED File Format
Large File Formats
Pairwise Interactions File Forma
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 Enser
Use my own data
Glossary
Supported browsers
Browse News by Topic
Archives
Archives: Table of assemblies
Search documentation...
Go
Help & Documentation
Using this website
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:
Human
Feature type:
Genes
Genomic locations
Variants
Identifier:
rs1333049
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
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
If you have a question that is not answered here, please contact our HelpDesk

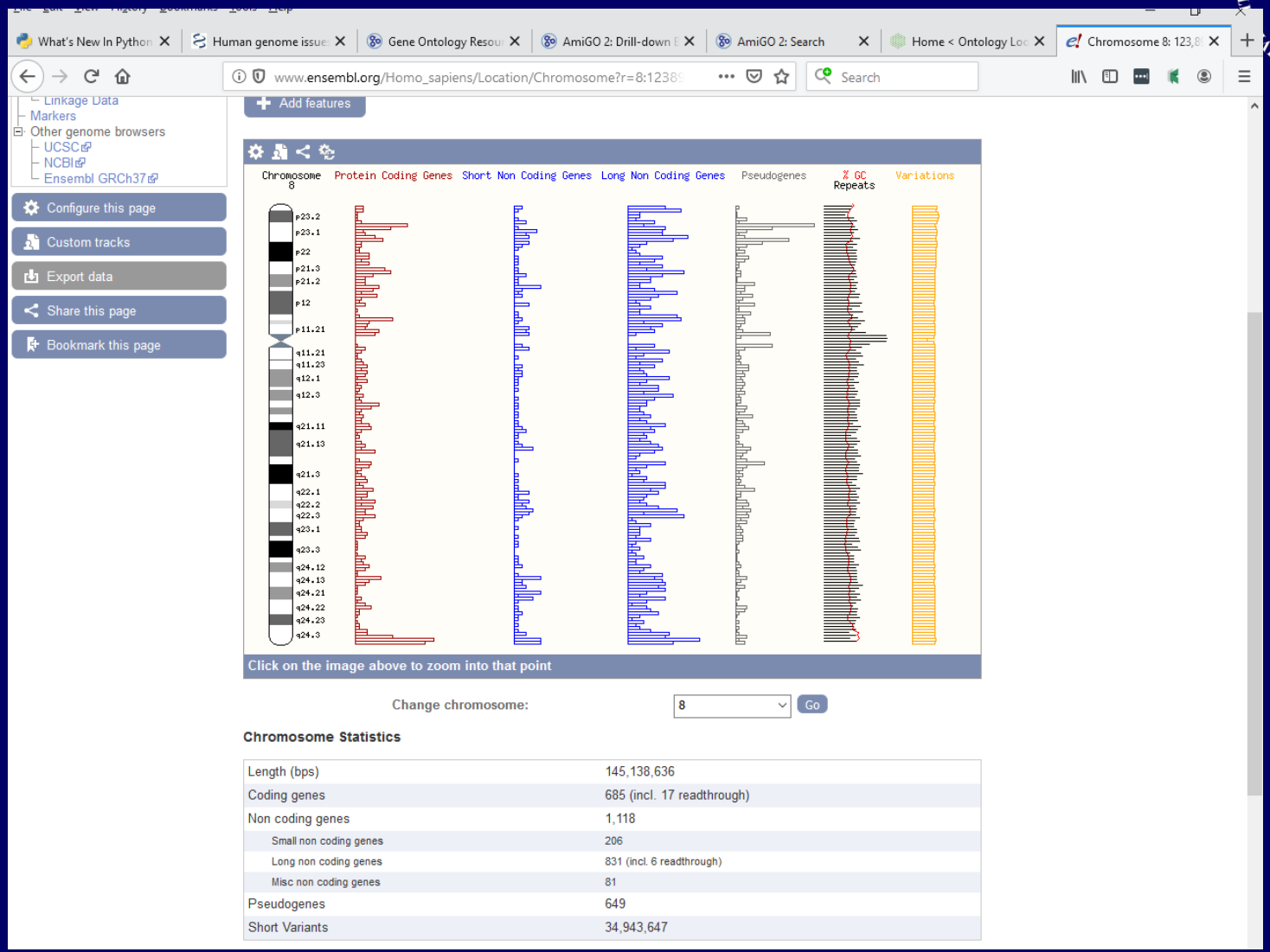

Ensembl geno...
Human Genom...
AmiGO 2: Drill-...
AmiGO 2: Search
Semmelweis Eg...
GO_0085014 < ...
GO_0085014 < ...
Annotation &a...
www.ensembl.org/info/genome/index.html
Search
Ensembl
BLAST/BLAT | BioMart | Tools | Downloads | Help & Documentation | Blog | Mirrors
Login/Register
Search all species...
Using this website
Annotation and prediction
Data access
API & software
About us
In this section
Ensembl stable IDs
Stable ID Version Increments in
Variation
Data Description
Predicted Data
Detailed species data count
Import VCF script
Phenotype Sources
Variant Sources
Comparative Genomics
EPO Pipeline
Gene Families
GeneTree stable ID
ncRNA trees and homologies
Orthology quality-controls
Protein trees and orthologies
Statistics on ncRNA trees
Statistics on protein trees
TreeFam HMM library
Whole-genome alignments
Microarray Probeset Mapping
Regulation
Regulatory Build
Regulation Sources
Regulatory Segmentation
Known Bugs
Search documentation...
Go
Help & Documentation
Annotation & Prediction
Annotation & Prediction
Ensembl creates automated annotation on a selection of chordate genomes, and also imports non-vertebrate model organisms for comparative purposes.
List of current species
In addition, a number of as-yet-unannotated vertebrate genomes are available on our pre! site, whilst our sister project Ensembl Genomes provides databases of bacteria, fungi, plants, metazoa and other non-vertebrates.
Ensembl annotation
Protein-coding genes are automatically annotated using Ensembl's genebuild pipeline. All transcripts are based on mRNA and proteins in public scientific databases.
Genome assemblies
Pax6
INS
FOXP2
BRCA2
DMD
ssh
Variation
The Ensembl Variation database stores areas of the genome that differ between individual genomes ("variants") and, where available, associated disease and phenotype information.
Introduction to variation data
Variant Effect Predictor (VEP)
ATCGAGCT
ATCCAGCT
ATCGAGAT
Comparative genomics
Ensembl focuses on two main areas of comparative genomics: creation of gene trees using representative proteins from each gene in a species, and alignment of DNA sequences to infer synteny, conservation, etc.
Protein trees and orthologies
Genomic alignments
Regulation
Ensembl Regulation provides resources describing the mechanisms of gene regulation in human and mouse cells. In particular, it focuses on transcriptional and post-transcriptional mechanisms.
Introduction to regulation data
Regulatory build
Ensembl release 86 - Oct 2016 © WTSI / EMBL-EBI
Permanent link
About Us
About us
Contact us
Citing Ensembl
Privacy policy
Get help
Using this website
Adding custom tracks
Downloading data
Video tutorials
Our sister sites
Ensembl Bacteria
Ensembl Fungi
Ensembl Plants
Ensembl Protists
Follow us
Blog
Twitter
Facebook

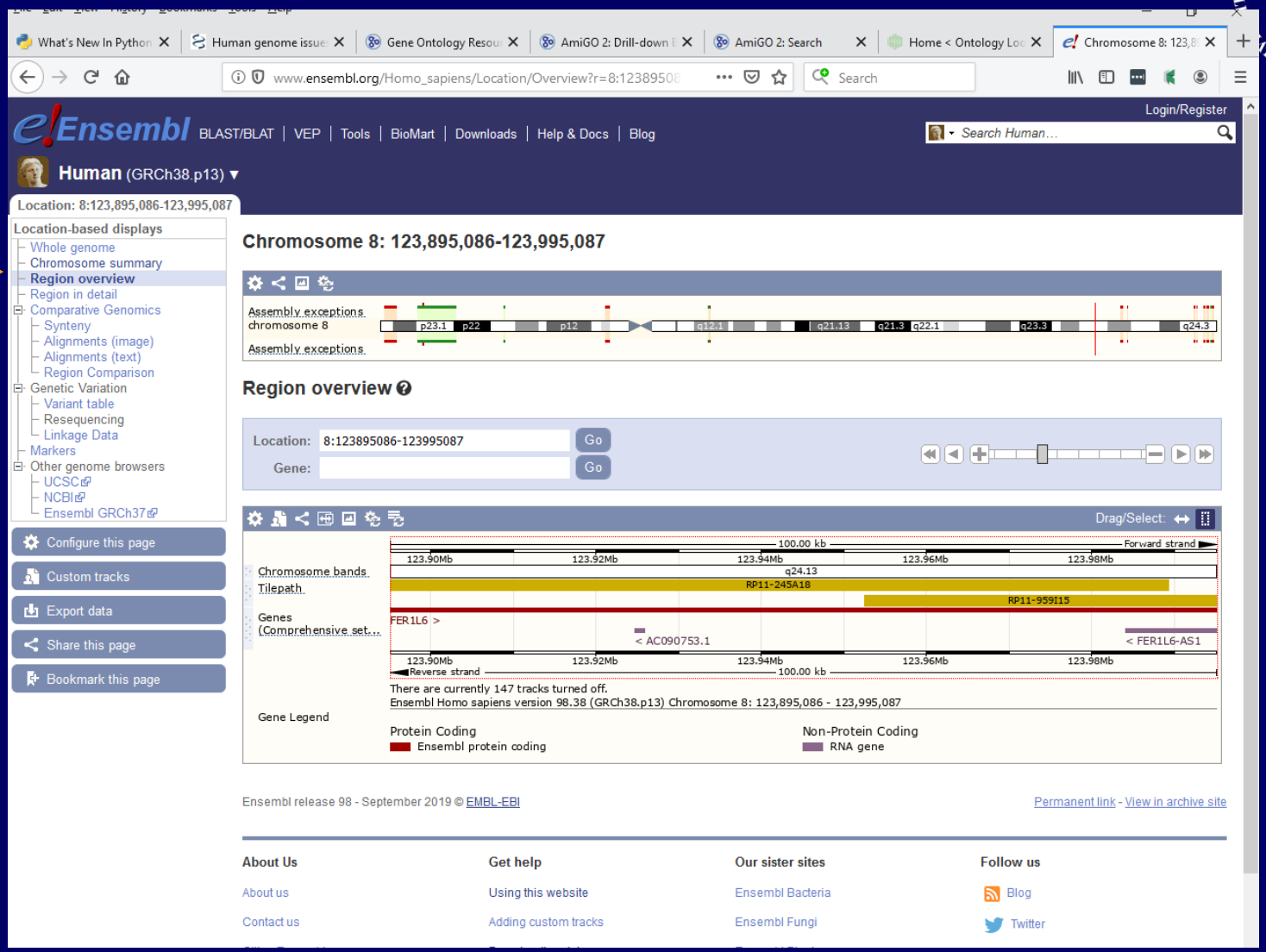

What's New In Python
Human genome issue
Gene Ontology Resou
AmiGO 2: Drill-down
AmiGO 2: Search
Home < Ontology Loc
Homo sapiens - Ensen
www.ensembl.org/Homo_sapiens/Info/Index
Search
Login/Register
Ensembl
BLAST/BLAT | VEP | Tools | BioMart | Downloads | Help & Docs | Blog
Search Human...
Human (GRCh38.p13)
Search Human (Homo sapiens)
Search all categories
Search Human...
Go
e.g. BRCA2 or 17:63992802-64038237 or rs699 or osteoarthritis
Genome assembly: GRCh38.p13 (GCA_000001405.28)
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
Go
View karyotype
Example region
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
Download FASTA files for genes, cDNAs, ncRNA, proteins
Download GTF or GFF3 files for genes, cDNAs, ncRNA, proteins
Update your old Ensembl IDs
Example gene
Example transcript
Comparative genomics
What can I find? Homologues, gene trees, and whole genome alignments across multiple species.
More about comparative analysis
Download alignments (EMF)
Example gene tree
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
Example variant
Example phenotype
Example structural variant
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)
Example regulatory feature
ENCODE

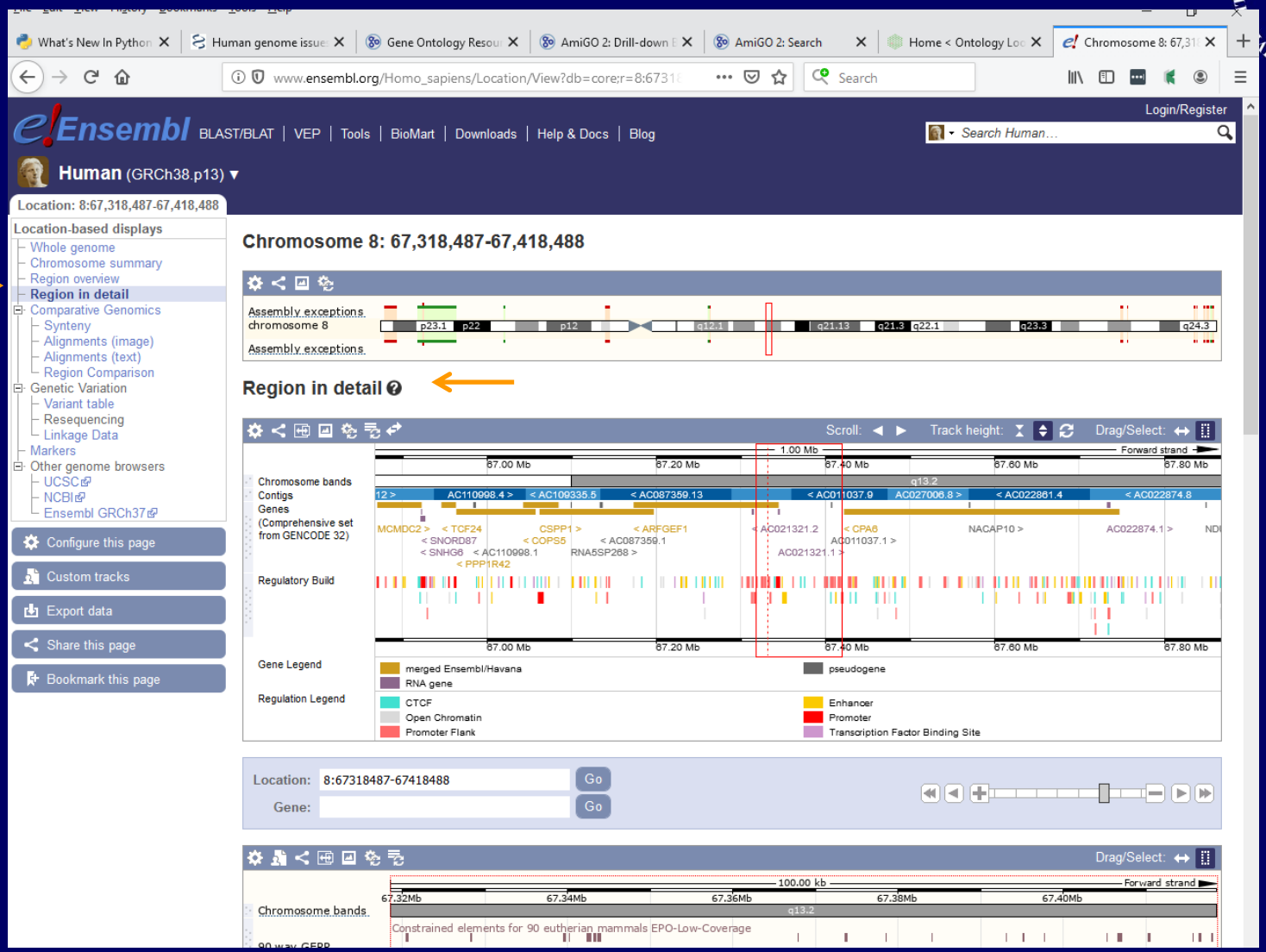

What's New In Python
Human genome issues
Gene Ontology Resou
AmiGO 2: Drill-down
AmiGO 2: Search
Home < Ontology Loc
Homo sapiens - Ensem
www.ensembl.org/Homo_sapiens/Info/Annotation
Search
Login/Register
Ensembl
BLAST/BLAT | VEP | Tools | BioMart | Downloads | Help & Docs | Blog
Search Human…
Human (GRCh38.p13)
Human assembly and gene annotation
Assembly
This site provides a data set based on the December 2013 Homo sapiens high coverage assembly GRCh38 from the Genome Reference Consortium. This assembly is used by UCSC to create their hg38 database. The data set consists of gene models built from the genewise alignments of the human proteome as well as from alignments of human cDNAs using the cDNA2genome model of exonerate.
This release of the assembly has the following properties:
contig length total 3.4 Gb.
chromosome length total 3.1 Gb (excluding haplotypes).
It also includes 261 alt loci scaffolds, mainly in the LRC/KIR complex on chromosome 19 (35 alternate sequence representations) and the MHC region on chromosome 6 (7 alternate sequence representations).
Watch a video on YouTube about patches and haplotypes in the Human genome.
Patches
As the GRC maintains and improves the assembly, patches are being introduced. Currently, assembly patches are of two types:
Novel patch: new sequences that add alternative sequence at a loci and will remain as haplotypes in the next major assembly release by GRC
Fix patch: sequences that correct the reference sequence and will replace the given region of the reference assembly at the next major assembly release by GRC.
The genome assembly represented here corresponds to GenBank Assembly ID GCA_000001405.28
Other assemblies
GRCh37 Full Feb 2014 archive with BLAST, VEP and BioMart
Go
Gene annotation
The Ensembl human gene annotations have been updated using Ensembl's automatic annotation pipeline. The updated annotation incorporates new protein and cDNA sequences which have become publicly available since the last GRCh38 genebuild (December 2013).
In the current release, we continue to display a joint gene set based on the merge between the automatic annotation from Ensembl and the manually curated annotation from Havana. See the statistics table,
Statistics
Summary
Assembly | GRCh38.p13 (Genome Reference Consortium Human Build 38), INSDC Assembly GCA_000001405.28, Dec 2013
Base Pairs | 4,537,931,177
Golden Path Length | 3,096,649,726
Annotation provider | Ensembl
Annotation method | Full genebuild
Genebuild started | Jan 2014
Genebuild released | Jul 2014
Genebuild last updated/patched | Jun 2019
Database version | 98.38
Gencode version | GENCODE 32
Gene counts (Primary assembly)
Coding genes | 20,444 (incl 667 readthrough)
Non coding genes | 23,949
Small non coding genes | 4,871
Long non coding genes | 16,857 (incl 304 readthrough)
Misc non coding genes | 2,221
Pseudogenes | 15,214 (incl 8 readthrough)
Gene transcripts | 227,530
Gene counts (Alternative sequence)
Coding genes | 3,054 (incl 44 readthrough)
Non coding genes | 1,555
Small non coding genes | 297
Long non coding genes | 1,071 (incl 45 readthrough)


What's New In Python
Human genome issues
Gene Ontology Resou
AmiGO 2: Drill-down
AmiGO 2: Search
Home < Ontology Loc
Whole genome - Hom
www.ensembl.org/Homo_sapiens/Location/Genome
Search
Ensembl BLAST/BLAT | VEP | Tools | BioMart | Downloads | Help & Docs | Blog
Login/Register
Search Human...
Human (GRCh38.p13)
Genome
Location-based displays
Whole genome
Chromosome summary
Region overview
Region in detail
Comparative Genomics
Synteny
Alignments (image)
Alignments (text)
Region Comparison
Genetic Variation
Variant table
Resequencing
Strain table
Linkage Data
Markers
Other genome browsers
UCSC
NCBI
Ensembl GRCh37
Configure this page
Custom tracks
Export data
Share this page
Bookmark this page
Whole genome
Add features
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 X Y MT
Click on the image above to jump to a chromosome, or click and drag to select a region
Summary
Assembly
GRCh38.p13 (Genome Reference Consortium Human Build 38), INSDC Assembly GCA_000001405.28, Dec 2013
Base Pairs
4,537,931,177
Golden Path Length
3,096,649,726
Annotation provider
Ensembl
Annotation method
Full genebuild
Genebuild started
Jan 2014
Genebuild released
Jul 2014
Genebuild last updated/patched
Jun 2019
Database version
98.38


What's New In Python
Human genome issue
Gene Ontology Resou
AmiGO 2: Drill-down
AmiGO 2: Search
Home < Ontology Loc
Chromosome 8: 123,8
www.ensembl.org/Homo_sapiens/Location/Chromosome?r=8:12385
Search
Login/Register
Ensembl BLAST/BLAT | VEP | Tools | BioMart | Downloads | Help & Docs | Blog
Search Human...
Human (GRCh38.p13)
Location: 8:123,895,086-123,995,087
Location-based displays
Whole genome
Chromosome summary
Region overview
Region in detail
Comparative Genomics
Synteny
Alignments (image)
Alignments (text)
Region Comparison
Genetic Variation
Variant table
Resequencing
Linkage Data
Markers
Other genome browsers
UCSC
NCBI
Ensembl GRCh37
Configure this page
Custom tracks
Export data
Share this page
Bookmark this page
Chromosome 8: 123,895,086-123,995,087
Assembly exceptions
chromosome 8
Assembly exceptions
Chromosome summary
Add features
Chromosome
Protein Coding Genes
Short Non Coding Genes
Long Non Coding Genes
Pseudogenes
% GC
Repeats
Variations


What's New In Python
Human genome issues
Gene Ontology Resou
AmiGO 2: Drill-down
AmiGO 2: Search
Home < Ontology Loc
Chromosome 8: 123,8
www.ensembl.org/Homo_sapiens/Location/Chromosome?r=8:12385
Search
Linkage Data
Markers
Other genome browsers
UCSC
NCBI
Ensembl GRCh37
Configure this page
Custom tracks
Export data
Share this page
Bookmark this page
Add features
Chromosome 8
Protein Coding Genes
Short Non Coding Genes
Long Non Coding Genes
Pseudogenes
% GC Repeats
Variations
p23.2
p23.1
p22
p21.3
p21.2
p12
p11.21
q11.21
q11.23
q12.1
q12.3
q21.11
q21.13
q21.3
q22.1
q22.2
q22.3
q23.1
q23.3
q24.12
q24.13
q24.21
q24.22
q24.23
q24.3
Click on the image above to zoom into that point
Change chromosome:
8
Go
Chromosome Statistics
Length (bps) 145,138,636
Coding genes 685 (incl. 17 readthrough)
Non coding genes 1,118
Small non coding genes 206
Long non coding genes 831 (incl. 6 readthrough)
Misc non coding genes 81
Pseudogenes 649
Short Variants 34,943,647


What's New In Python
Human genome issues
Gene Ontology Resou
AmiGO 2: Drill-down
AmiGO 2: Search
Home < Ontology Loc
Chromosome 8: 123,8
www.ensembl.org/Homo_sapiens/Location/Overview?r=8:12389508
Search
Ensembl BLAST/BLAT | VEP | Tools | BioMart | Downloads | Help & Docs | Blog
Login/Register
Search Human...
Human (GRCh38.p13)
Location: 8:123,895,086-123,995,087
Location-based displays
Whole genome
Chromosome summary
Region overview
Region in detail
Comparative Genomics
Synteny
Alignments (image)
Alignments (text)
Region Comparison
Genetic Variation
Variant table
Resequencing
Linkage Data
Markers
Other genome browsers
UCSC
NCBI
Ensembl GRCh37
Configure this page
Custom tracks
Export data
Share this page
Bookmark this page
Chromosome 8: 123,895,086-123,995,087
Assembly exceptions
chromosome 8
Assembly exceptions
Region overview
Location: 8:123895086-123995087
Go
Gene:
Go
Drag/Select:
Chromosome bands
Tilepath
Genes
(Comprehensive set...
100.00 kb
Forward strand
123.90Mb
123.92Mb
123.94Mb
123.96Mb
123.98Mb
q24.13
RP11-245A18
RP11-959I15
FER1L6 >
< AC090753.1
< FER1L6-AS1
Reverse strand
There are currently 147 tracks turned off.
Ensembl Homo sapiens version 98.38 (GRCh38.p13) Chromosome 8: 123,895,086 - 123,995,087
Gene Legend
Protein Coding
Ensembl protein coding
Non-Protein Coding
RNA gene
Ensembl release 98 - September 2019 © EMBL-EBI
Permanent link - View in archive site
About Us
About us
Contact us
Get help
Using this website
Adding custom tracks
Our sister sites
Ensembl Bacteria
Ensembl Fungi
Follow us
Blog
Twitter


Ensembl BLAST/BLAT | VEP | Tools | BioMart | Downloads | Help & Docs | Blog
Human (GRCh38.p13)
Location: 8:67,318,487-67,418,488
Location-based displays
Whole genome
Chromosome summary
Region overview
Region in detail
Comparative Genomics
Synteny
Alignments (image)
Alignments (text)
Region Comparison
Genetic Variation
Variant table
Resequencing
Linkage Data
Markers
Other genome browsers
UCSC
NCBI
Ensembl GRCh37
Configure this page
Custom tracks
Export data
Share this page
Bookmark this page
Chromosome 8: 67,318,487-67,418,488
Region in detail
Chromosome bands
Contigs
Genes
(Comprehensive set from GENCODE 32)
Regulatory Build
Gene Legend
merged Ensembl/Havana
RNA gene
pseudogene
Regulation Legend
CTCF
Open Chromatin
Promoter Flank
Enhancer
Promoter
Transcription Factor Binding Site
Location: 8:67318487-67418488
Gene:
Go
Constrained elements for 90 eutherian mammals EPO-Low-Coverage

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

Ensembl

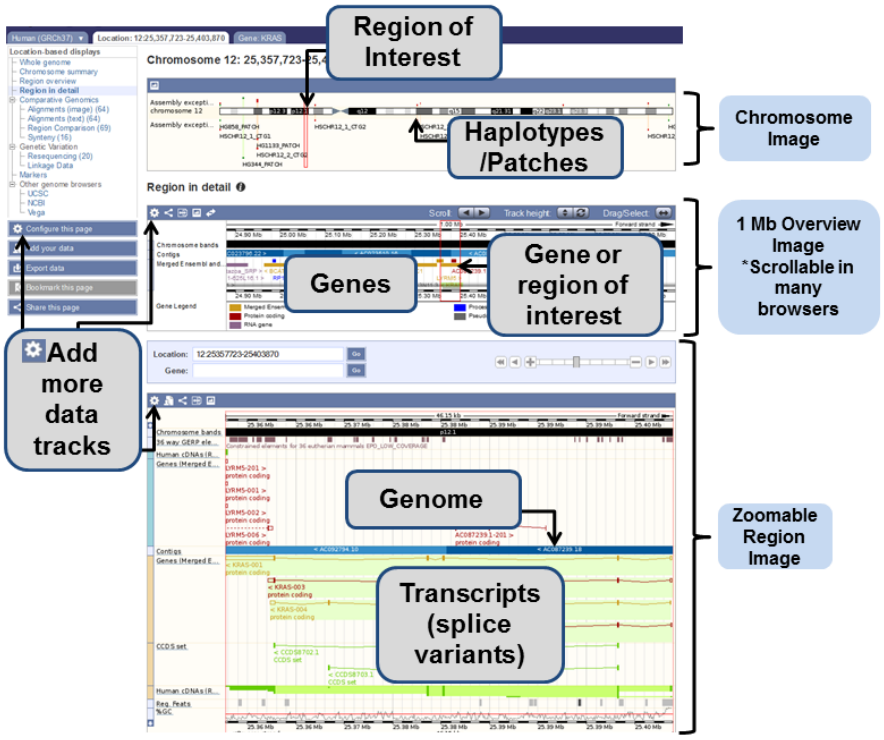
Help
- Search
- Help topics
  - Frequently Asked Questions
  - Video Tutorials
  - Glossary
- Contact HelpDesk

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



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

saved configuration

Reset configuration

Reset track order

Add your data

Go back to default settings

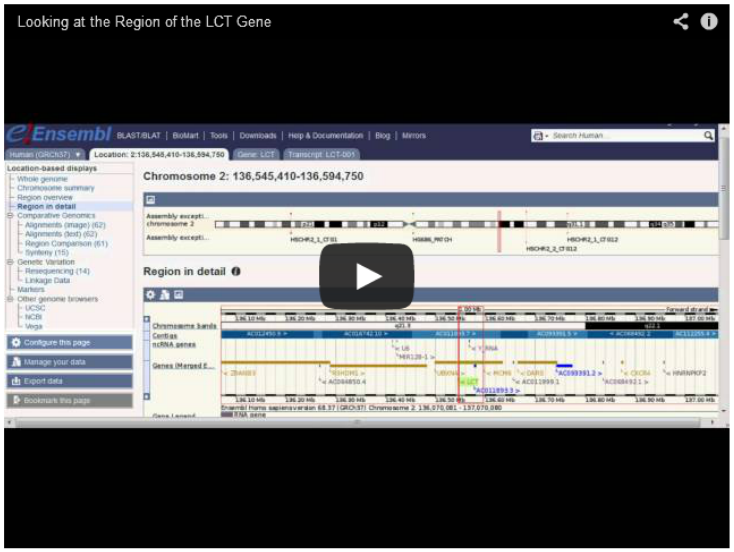
Input your own data

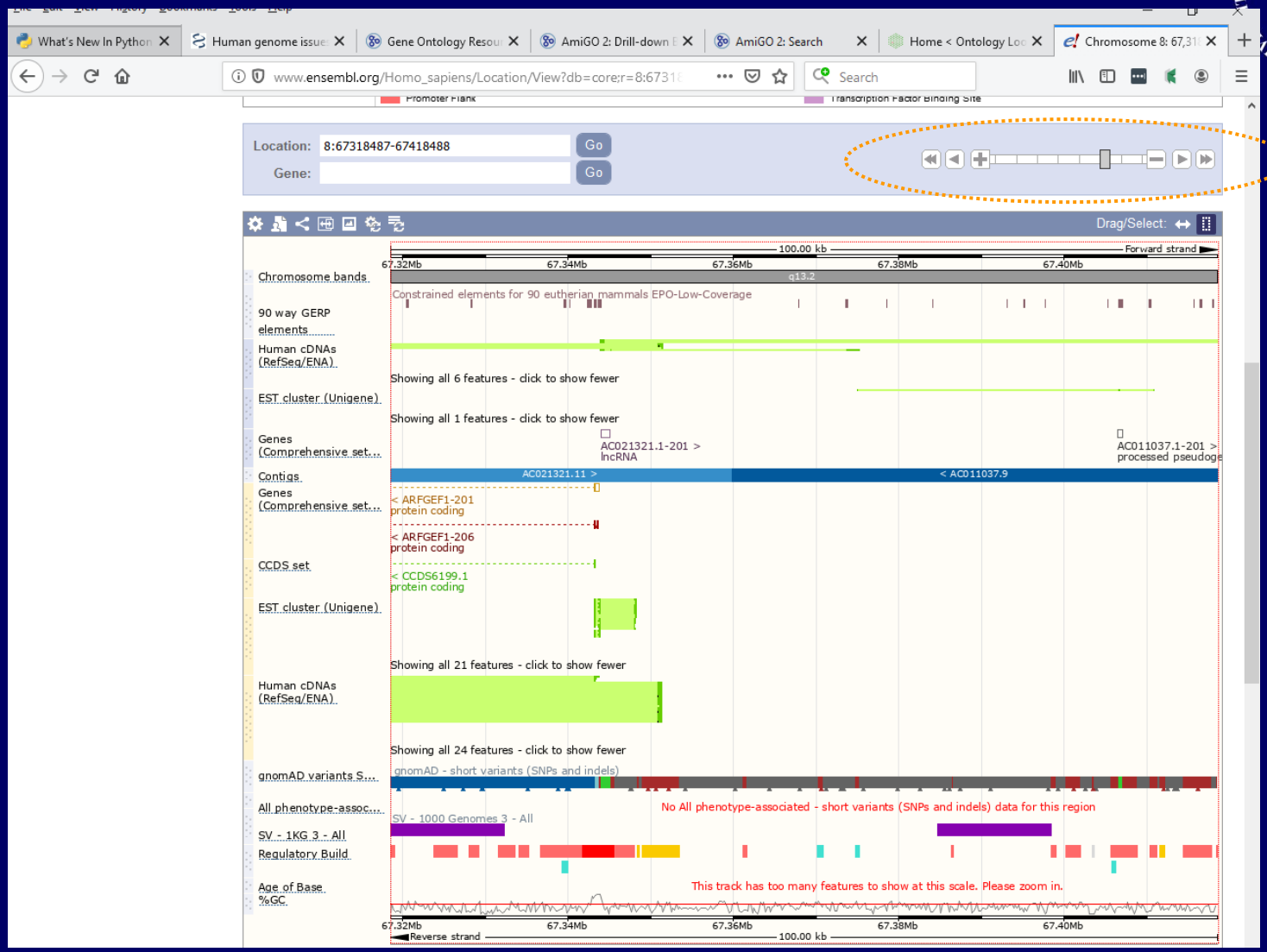
Add or change data tracks using the **configure this page** tool button at the left of the page, the cog wheel icons in the images, or click on the track name itself. Add data by navigating to the **Custom data** tab in the configuration window and clicking on **Add a custom track.**

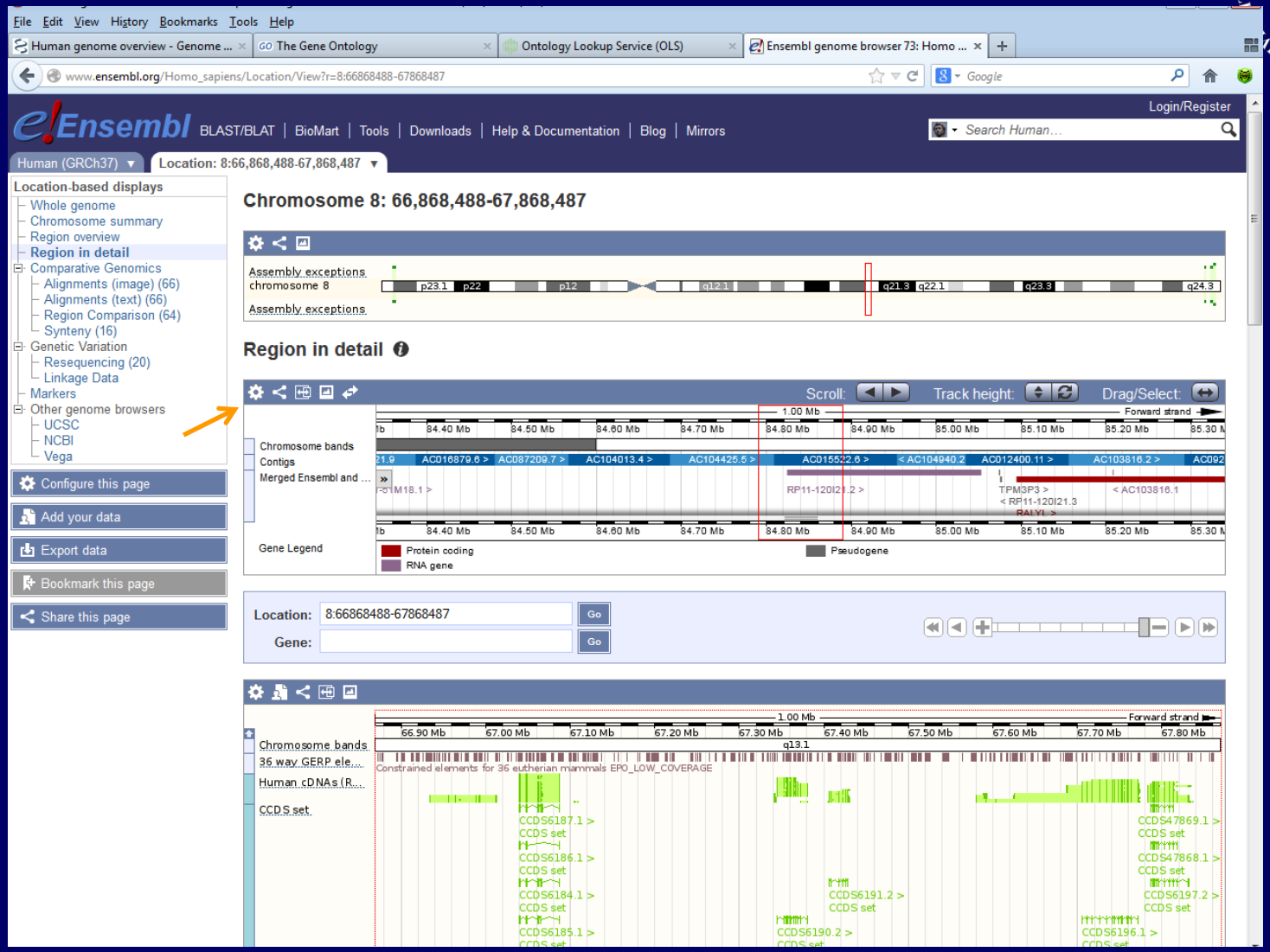
You can also attach your own data eg. BAM files by clicking the **Add your data button** (or **Manage your data** if you already have custom data added) and uploading or attaching a file, from the main view.

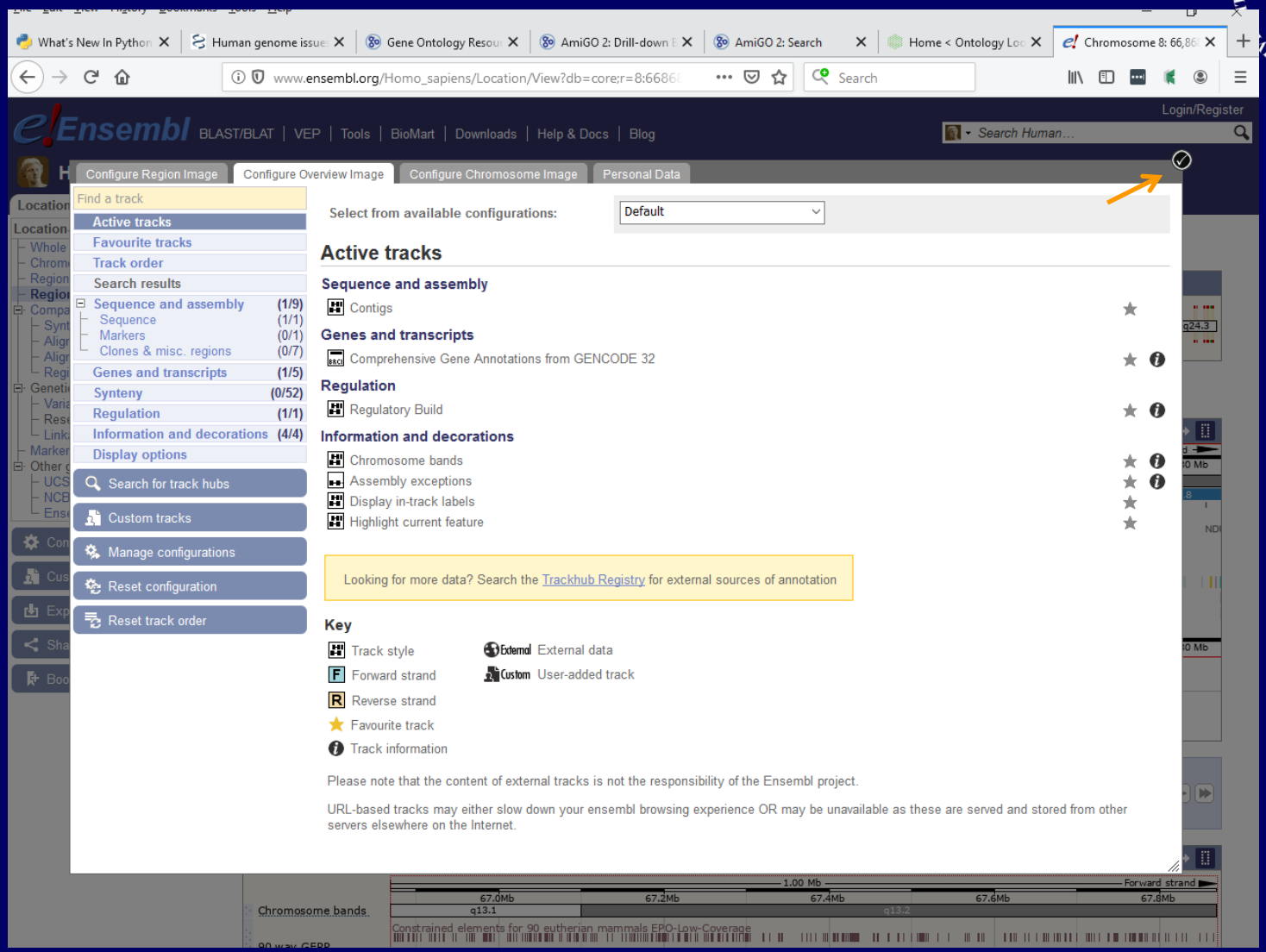
Some tracks can be displayed in different styles, however there is a limit on how much data can be displayed in certain styles. This FAQ explains the styles available and the amount of data that can be displayed.

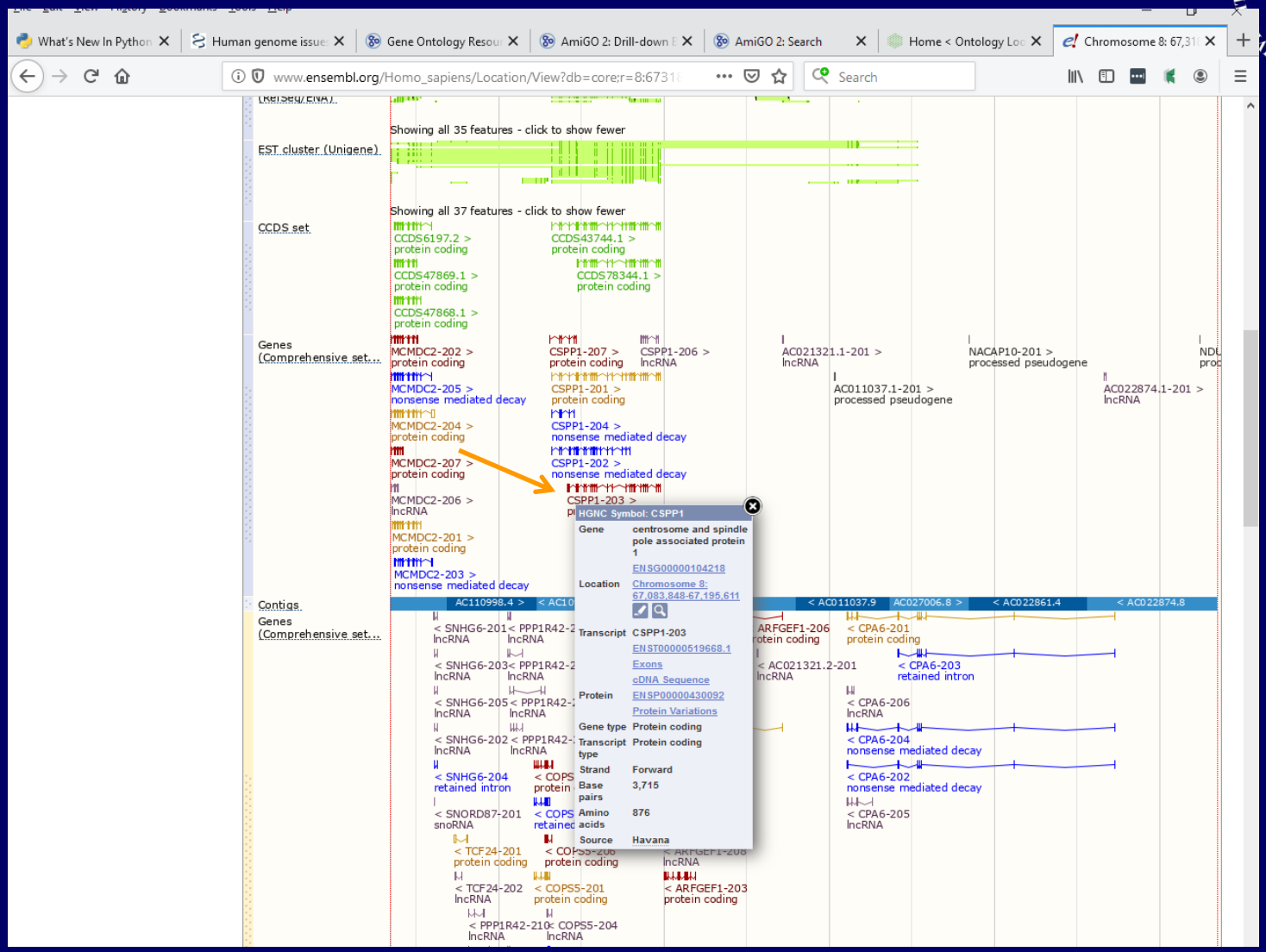
Looking at the Region of the LCT Gene - The Region in Detail View

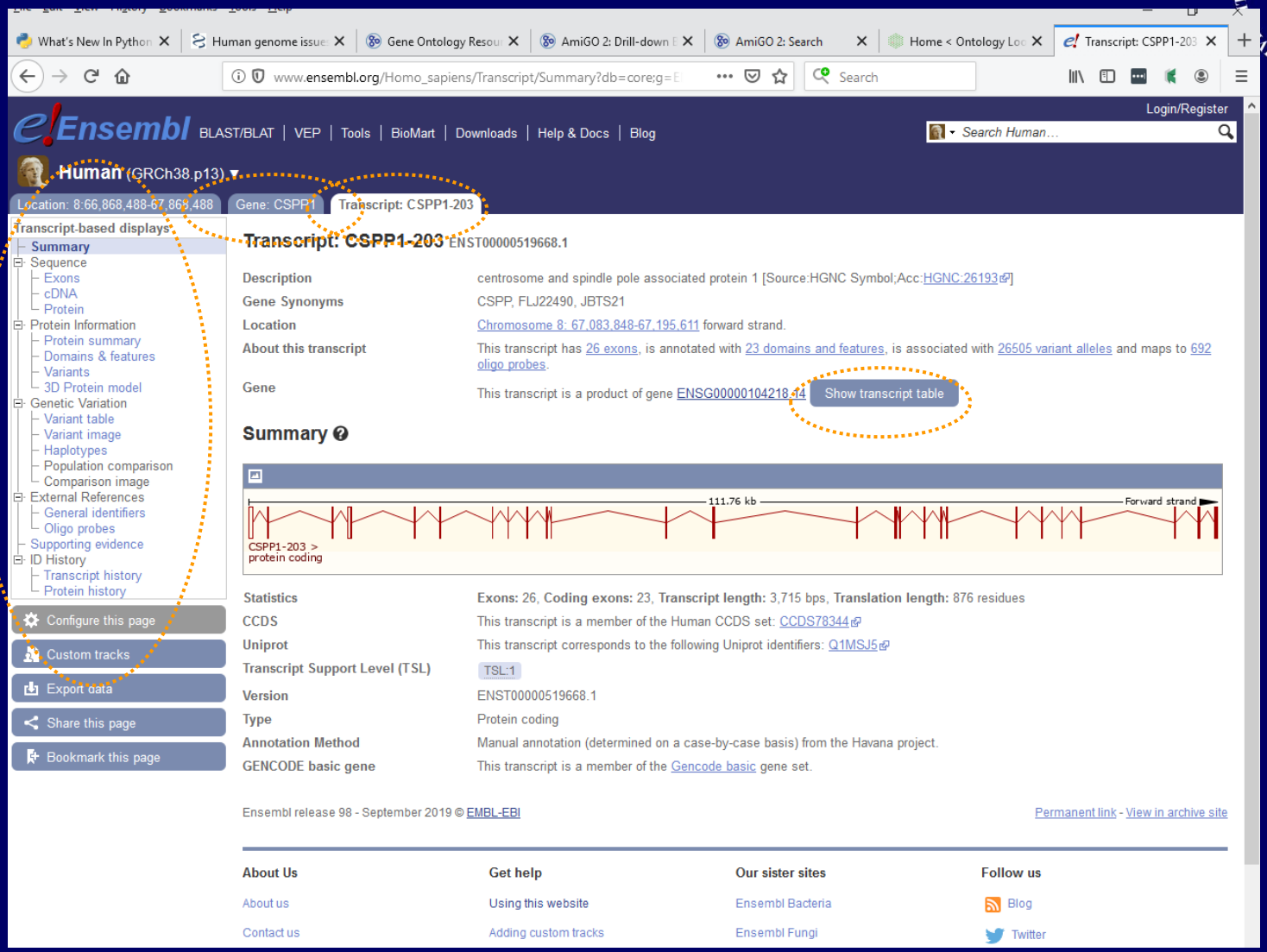


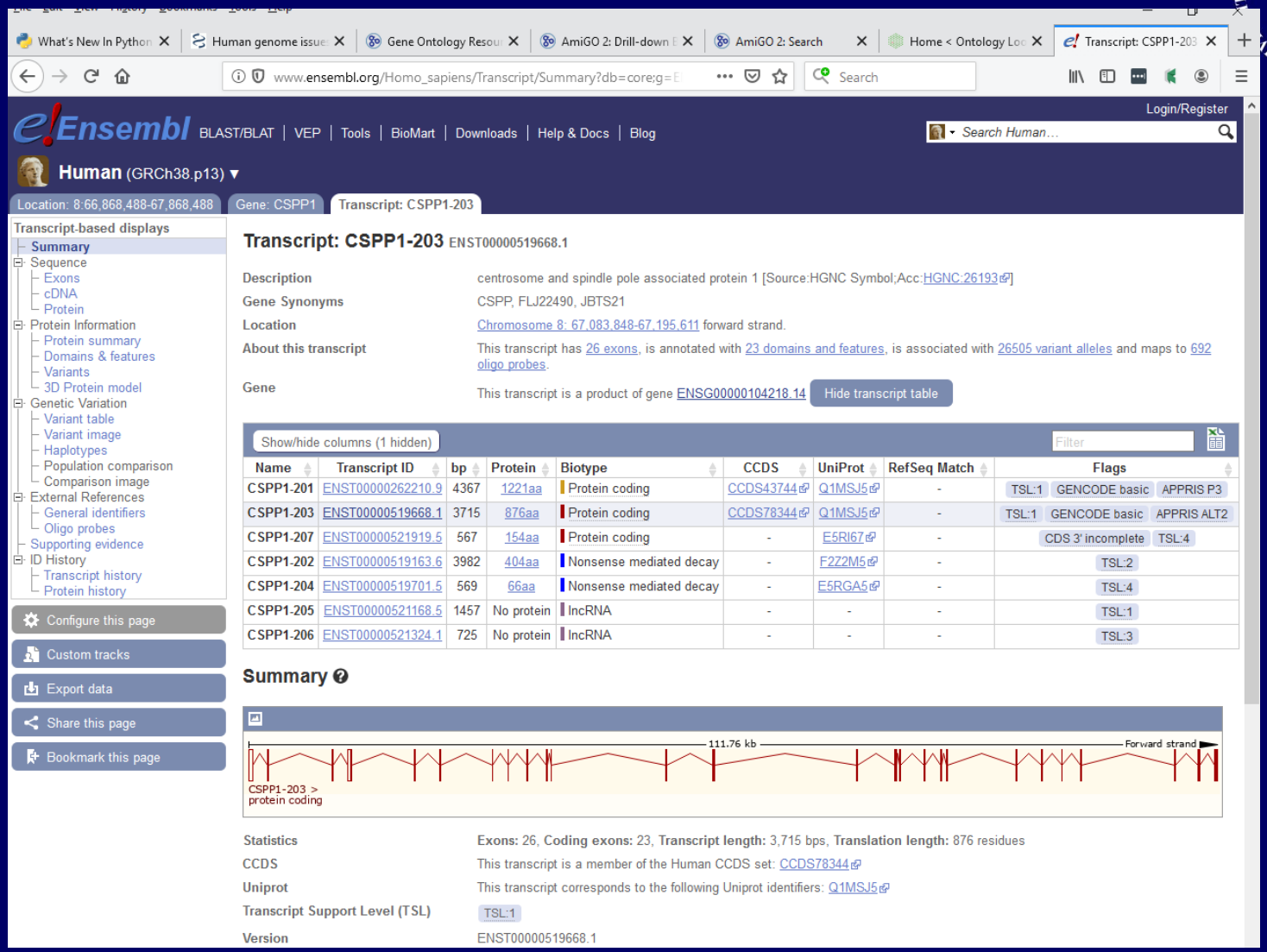

What's New In Python
Human genome issues
Gene Ontology Resou
AmiGO 2: Drill-down
AmiGO 2: Search
Home < Ontology Loc
Chromosome 8: 67,31
www.ensembl.org/Homo_sapiens/Location/View?db=core;r=8:67318
Search
Location: 8:67318487-67418488
Go
Gene:
Go
Drag/Select:
100.00 kb
Forward strand
67.32Mb
67.34Mb
67.36Mb
67.38Mb
67.40Mb
Chromosome bands
q13.2
Constrained elements for 90 eutherian mammals EPO-Low-Coverage
90 way GERP elements
Human cDNAs (RefSeq/ENA)
Showing all 6 features - click to show fewer
EST cluster (Unigene)
Showing all 1 features - click to show fewer
Genes (Comprehensive set...
AC021321.1-201 >
lncRNA
AC011037.1-201 >
processed pseudoge
Contigs
AC021321.11 >
< AC011037.9
Genes (Comprehensive set...
< ARFGEF1-201
protein coding
< ARFGEF1-206
protein coding
CCDS set
< CCDS6199.1
protein coding
EST cluster (Unigene)
Showing all 21 features - click to show fewer
Human cDNAs (RefSeq/ENA)
Showing all 24 features - click to show fewer
gnomAD variants S...
gnomAD - short variants (SNPs and indels)
All phenotype-assoc...
No All phenotype-associated - short variants (SNPs and indels) data for this region
SV - 1000 Genomes 3 - All
SV - 1KG 3 - All
Regulatory Build
Age of Base %GC
This track has too many features to show at this scale. Please zoom in.
Reverse strand
100.00 kb

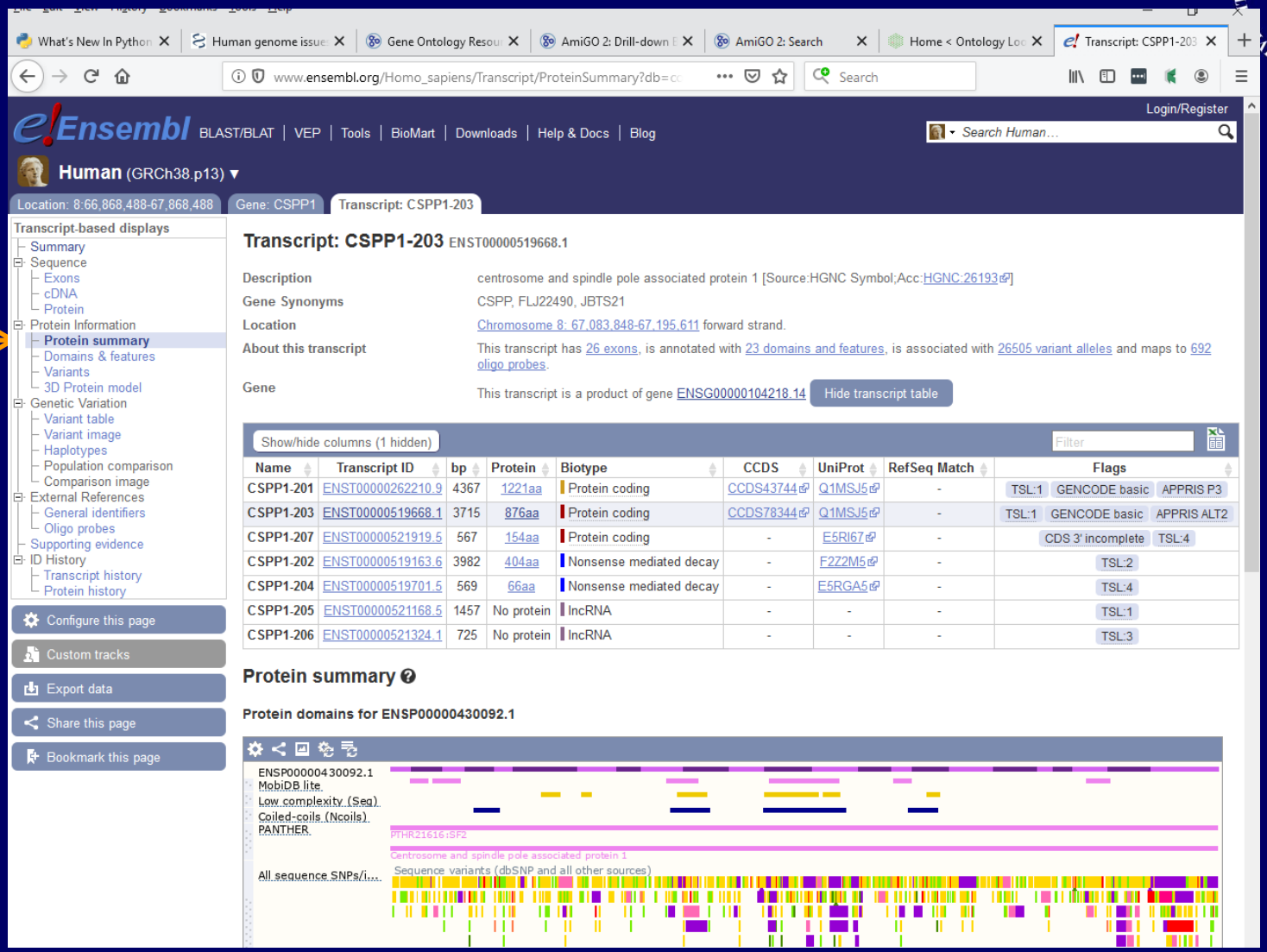

File Edit View History Bookmarks Tools Help
Human genome overview - Genome ...
The Gene Ontology
Ontology Lookup Service (OLS)
Ensembl genome browser 73: Homo ...
www.ensembl.org/Homo_sapiens/Location/View?r=8:66868488-67868487
Ensembl BLAST/BLAT | BioMart | Tools | Downloads | Help & Documentation | Blog | Mirrors
Login/Register
Search Human...
Human (GRCh37)
Location: 8:66,868,488-67,868,487
Location-based displays
Whole genome
Chromosome summary
Region overview
Region in detail
Comparative Genomics
Alignments (image) (66)
Alignments (text) (66)
Region Comparison (64)
Synteny (16)
Genetic Variation
Resequencing (20)
Linkage Data
Markers
Other genome browsers
UCSC
NCBI
Vega
Configure this page
Add your data
Export data
Bookmark this page
Share this page
Chromosome 8: 66,868,488-67,868,487
Assembly exceptions
chromosome 8
Region in detail
Scroll:
Track height:
Drag/Select:
Chromosome bands
Contigs
Merged Ensembl and ...
Gene Legend
Protein coding
RNA gene
Pseudogene
Location: 8:66868488-67868487
Gene:
Go
Constrained elements for 36 eutherian mammals EPO_LOW_COVERAGE
CCDS set

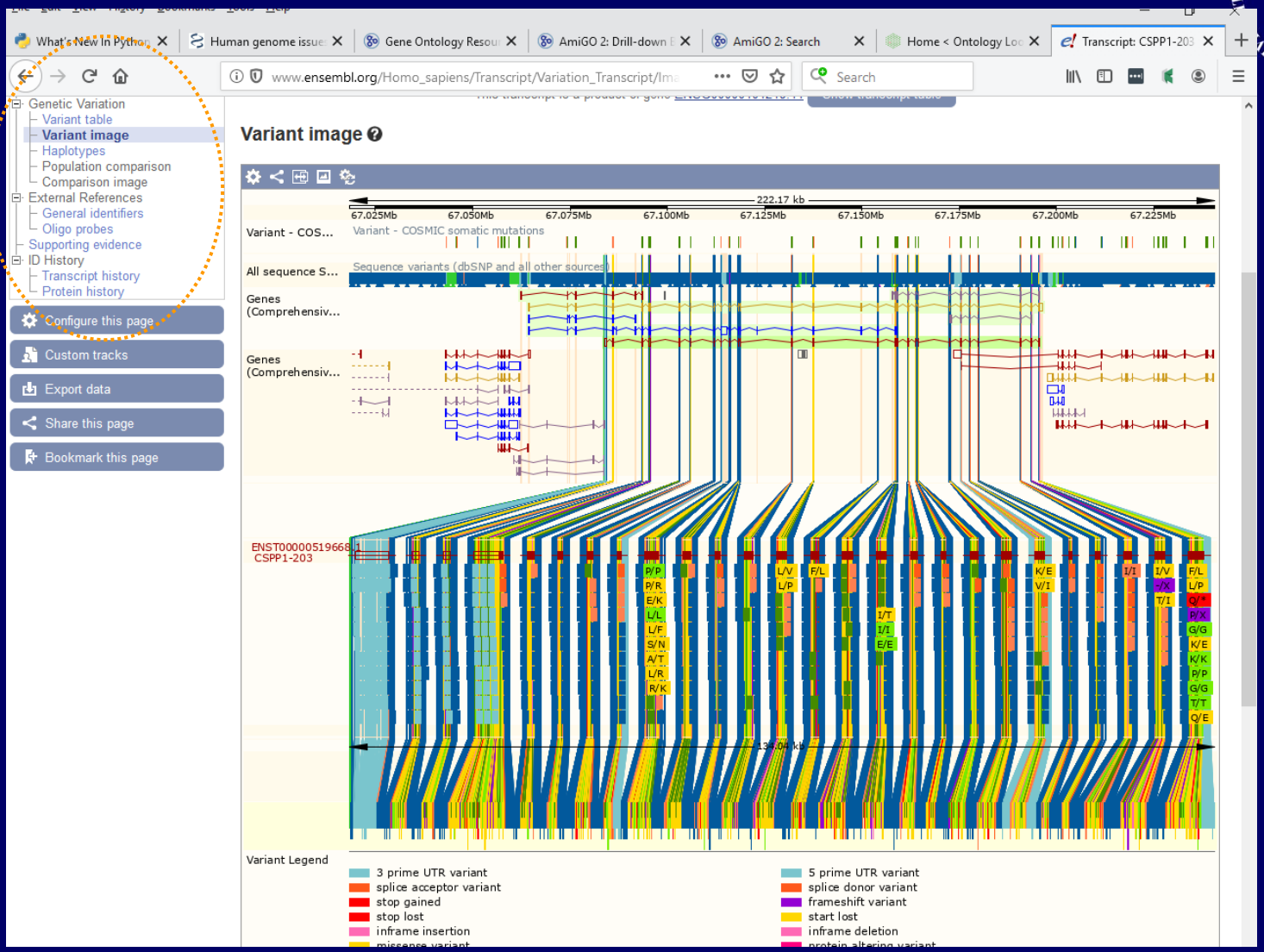

Ensembl
BLAST/BLAT | VEP | Tools | BioMart | Downloads | Help & Docs | Blog
Login/Register
Search Human...
Configure Region Image
Configure Overview Image
Configure Chromosome Image
Personal Data
Find a track
Active tracks
Favourite tracks
Track order
Search results
Sequence and assembly (1/9)
Sequence (1/1)
Markers (0/1)
Clones & misc. regions (0/7)
Genes and transcripts (1/5)
Synteny (0/52)
Regulation (1/1)
Information and decorations (4/4)
Display options
Search for track hubs
Custom tracks
Manage configurations
Reset configuration
Reset track order
Select from available configurations:
Default
Active tracks
Sequence and assembly
Contigs
Genes and transcripts
Comprehensive Gene Annotations from GENCODE 32
Regulation
Regulatory Build
Information and decorations
Chromosome bands
Assembly exceptions
Display in-track labels
Highlight current feature
Looking for more data? Search the Trackhub Registry for external sources of annotation
Key
Track style
Forward strand
Reverse strand
Favourite track
Track information
External data
User-added track
Please note that the content of external tracks is not the responsibility of the Ensembl project.
URL-based 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.
Chromosome bands
Constrained elements for 90 eutherian mammals EPO-Low-Coverage

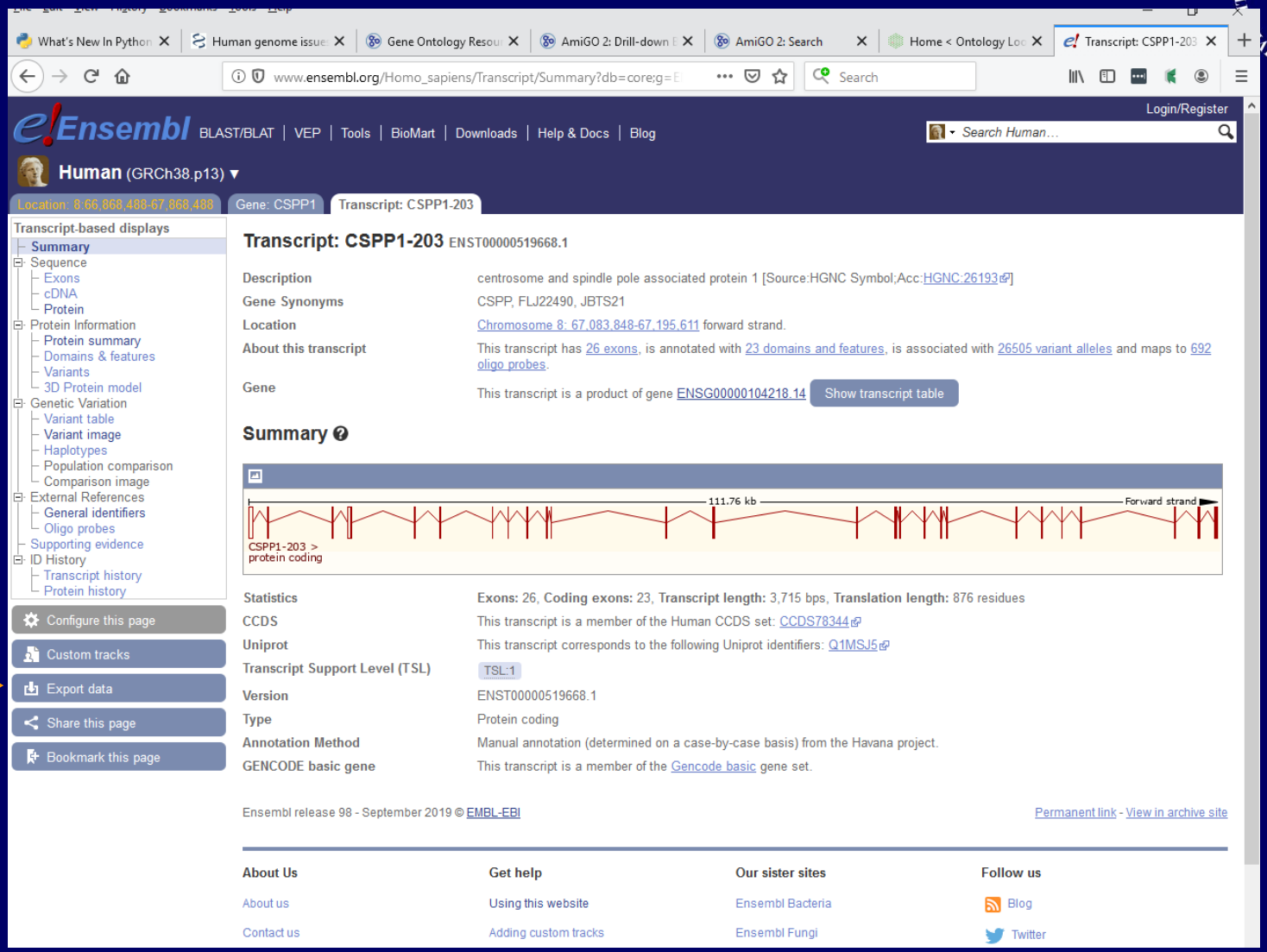

What's New In Python
Human genome issues
Gene Ontology Resou
AmiGO 2: Drill-down
AmiGO 2: Search
Home < Ontology Loc
Chromosome 8: 67,31
www.ensembl.org/Homo_sapiens/Location/View?db=core;r=8:67
Search
Showing all 35 features - click to show fewer
EST cluster (Unigene)
Showing all 37 features - click to show fewer
CCDS set
CCDS6197.2 > protein coding
CCDS43744.1 > protein coding
CCDS47869.1 > protein coding
CCDS78344.1 > protein coding
CCDS47868.1 > protein coding
Genes (Comprehensive set...
MCMDC2-202 > protein coding
CSPP1-207 > protein coding
CSPP1-206 > lncRNA
AC021321.1-201 > lncRNA
NACAP10-201 > processed pseudogene
AC011037.1-201 > processed pseudogene
AC022874.1-201 > lncRNA
MCMDC2-205 > nonsense mediated decay
CSPP1-201 > protein coding
MCMDC2-204 > protein coding
CSPP1-204 > nonsense mediated decay
MCMDC2-207 > protein coding
CSPP1-202 > nonsense mediated decay
MCMDC2-206 > lncRNA
CSPP1-203 >
MCMDC2-201 > protein coding
MCMDC2-203 > nonsense mediated decay
HGNC Symbol: CSPP1
Gene
centrosome and spindle pole associated protein 1
ENSG00000104218
Location
Chromosome 8: 67,083,848-67,195,611
Transcript
CSPP1-203
ENST00000519668.1
Exons
cDNA Sequence
Protein
ENSP00000430092
Protein Variations
Gene type
Protein coding
Transcript type
Protein coding
Strand
Forward
Base pairs
3,715
Amino acids
876
Source
Havana
Contigs
Genes (Comprehensive set...
AC110998.4 >
AC011037.9
AC027006.8 >
AC022861.4
AC022874.8
SNHG6-201 lncRNA
PPP1R42-2 lncRNA
SNHG6-203 lncRNA
SNHG6-205 lncRNA
SNHG6-202 lncRNA
SNHG6-204 retained intron
SNORD87-201 snoRNA
TCF24-201 protein coding
TCF24-202 lncRNA
PPP1R42-210 lncRNA
COPS5-201 protein coding
COPS5-204 lncRNA
ARFGEF1-206 protein coding
ARFGEF1-208 lncRNA
ARFGEF1-203 protein coding
AC021321.2-201 lncRNA
CPA6-201 protein coding
CPA6-203 retained intron
CPA6-206 lncRNA
CPA6-204 nonsense mediated decay
CPA6-202 nonsense mediated decay
CPA6-205 lncRNA

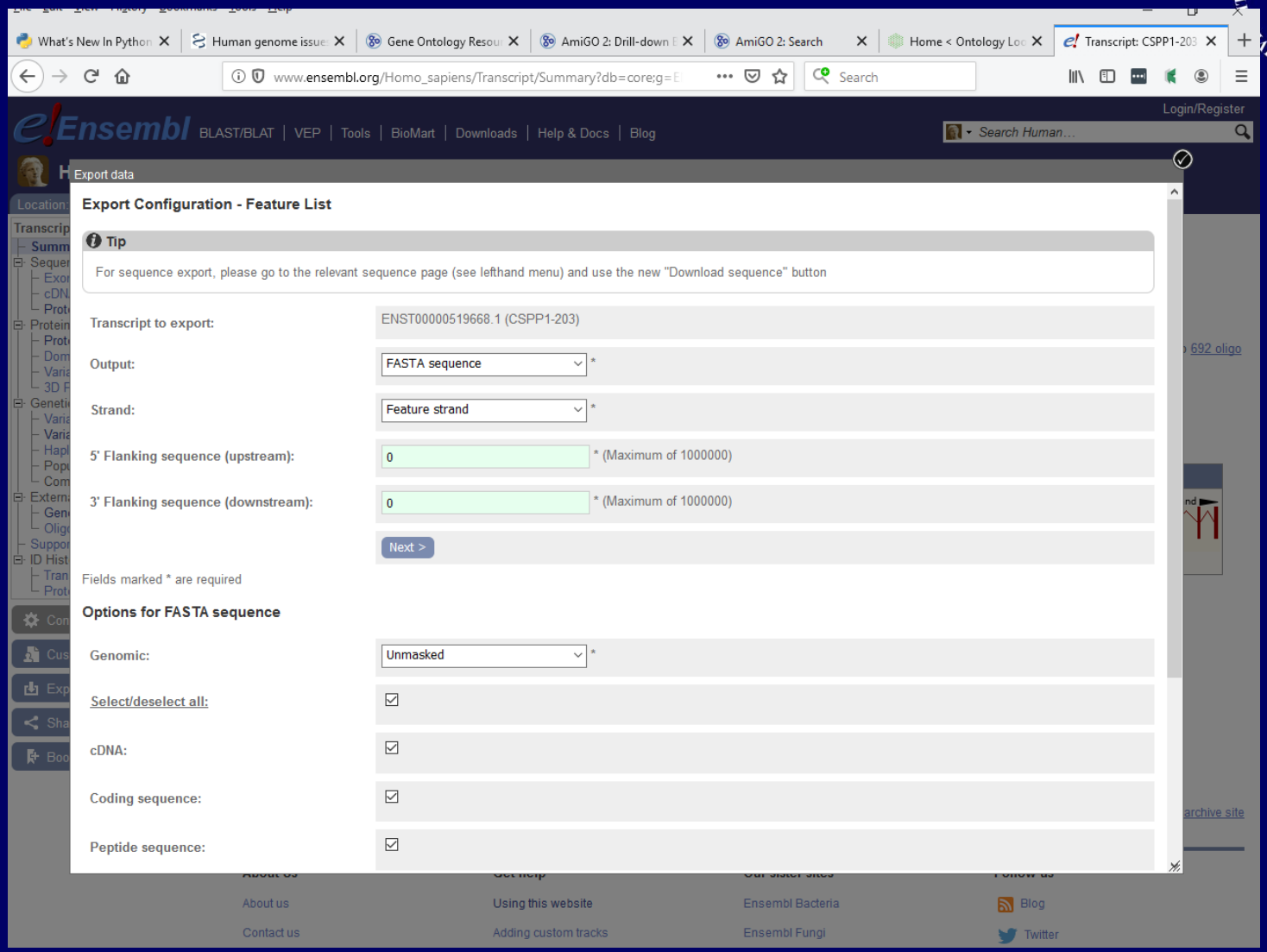

What's New In Python
Human genome issues
Gene Ontology Resou
AmiGO 2: Drill-down
AmiGO 2: Search
Home < Ontology Loc
Transcript: CSPP1-203
www.ensembl.org/Homo_sapiens/Transcript/Summary?db=core;g=E
Search
Ensembl
BLAST/BLAT | VEP | Tools | BioMart | Downloads | Help & Docs | Blog
Login/Register
Search Human...
Human (GRCh38.p13)
Location: 8:66,868,488-67,868,488
Gene: CSPP1
Transcript: CSPP1-203
Transcript-based displays
Summary
Sequence
Exons
cDNA
Protein
Protein Information
Protein summary
Domains & features
Variants
3D Protein model
Genetic Variation
Variant table
Variant image
Haplotypes
Population comparison
Comparison image
External References
General identifiers
Oligo probes
Supporting evidence
ID History
Transcript history
Protein history
Configure this page
Custom tracks
Export data
Share this page
Bookmark this page
Transcript: CSPP1-203 ENST00000519668.1
Description
centrosome and spindle pole associated protein 1 [Source:HGNC Symbol;Acc:HGNC:26193]
Gene Synonyms
CSPP, FLJ22490, JBTS21
Location
Chromosome 8: 67,083,848-67,195,611 forward strand.
About this transcript
This transcript has 26 exons, is annotated with 23 domains and features, is associated with 26505 variant alleles and maps to 692 oligo probes.
Gene
This transcript is a product of gene ENSG00000104218.14
Show transcript table
Summary
111.76 kb
Forward strand
CSPP1-203 >
protein coding
Statistics
Exons: 26, Coding exons: 23, Transcript length: 3,715 bps, Translation length: 876 residues
CCDS
This transcript is a member of the Human CCDS set: CCDS78344
Uniprot
This transcript corresponds to the following Uniprot identifiers: Q1MSJ5
Transcript Support Level (TSL)
TSL:1
Version
ENST00000519668.1
Type
Protein coding
Annotation Method
Manual annotation (determined on a case-by-case basis) from the Havana project.
GENCODE basic gene
This transcript is a member of the Gencode basic gene set.
Ensembl release 98 - September 2019 © EMBL-EBI
Permanent link - View in archive site
About Us
About us
Contact us
Get help
Using this website
Adding custom tracks
Our sister sites
Ensembl Bacteria
Ensembl Fungi
Follow us
Blog
Twitter


What's New In Python
Human genome issue
Gene Ontology Resou
AmiGO 2: Drill-down
AmiGO 2: Search
Home < Ontology Loc
Transcript: CSPP1-203
www.ensembl.org/Homo_sapiens/Transcript/Summary?db=core;g=E
Search
Ensembl
BLAST/BLAT | VEP | Tools | BioMart | Downloads | Help & Docs | Blog
Login/Register
Search Human...
Human (GRCh38.p13)
Location: 8:66,868,488-67,868,488
Gene: CSPP1
Transcript: CSPP1-203
Transcript-based displays
Summary
Sequence
Exons
cDNA
Protein
Protein Information
Protein summary
Domains & features
Variants
3D Protein model
Genetic Variation
Variant table
Variant image
Haplotypes
Population comparison
Comparison image
External References
General identifiers
Oligo probes
Supporting evidence
ID History
Transcript history
Protein history
Configure this page
Custom tracks
Export data
Share this page
Bookmark this page
Transcript: CSPP1-203 ENST00000519668.1
Description
centrosome and spindle pole associated protein 1 [Source:HGNC Symbol;Acc:HGNC:26193]
Gene Synonyms
CSPP, FLJ22490, JBTS21
Location
Chromosome 8: 67,083,848-67,195,611 forward strand.
About this transcript
This transcript has 26 exons, is annotated with 23 domains and features, is associated with 26505 variant alleles and maps to 692 oligo probes.
Gene
This transcript is a product of gene ENSG00000104218.14
Hide transcript table
Show/hide columns (1 hidden)
Filter
Name | Transcript ID | bp | Protein | Biotype | CCDS | UniProt | RefSeq Match | Flags
CSPP1-201 | ENST00000262210.9 | 4367 | 1221aa | Protein coding | CCDS43744 | Q1MSJ5 | - | TSL:1 GENCODE basic APPRIS P3
CSPP1-203 | ENST00000519668.1 | 3715 | 876aa | Protein coding | CCDS78344 | Q1MSJ5 | - | TSL:1 GENCODE basic APPRIS ALT2
CSPP1-207 | ENST00000521919.5 | 567 | 154aa | Protein coding | - | E5RI67 | - | CDS 3' incomplete TSL:4
CSPP1-202 | ENST00000519163.6 | 3982 | 404aa | Nonsense mediated decay | - | F2Z2M5 | - | TSL:2
CSPP1-204 | ENST00000519701.5 | 569 | 66aa | Nonsense mediated decay | - | E5RGA5 | - | TSL:4
CSPP1-205 | ENST00000521168.5 | 1457 | No protein | lncRNA | - | - | - | TSL:1
CSPP1-206 | ENST00000521324.1 | 725 | No protein | lncRNA | - | - | - | TSL:3
Summary
111.76 kb
Forward strand
CSPP1-203 > protein coding
Statistics
Exons: 26, Coding exons: 23, Transcript length: 3,715 bps, Translation length: 876 residues
CCDS
This transcript is a member of the Human CCDS set: CCDS78344
Uniprot
This transcript corresponds to the following Uniprot identifiers: Q1MSJ5
Transcript Support Level (TSL)
TSL:1
Version
ENST00000519668.1


What's New In Python
Human genome issues
Gene Ontology Resou
AmiGO 2: Drill-down
AmiGO 2: Search
Home < Ontology Loc
Transcript: CSPP1-203
www.ensembl.org/Homo_sapiens/Transcript/ProteinSummary?db=c
Search
Ensembl
BLAST/BLAT | VEP | Tools | BioMart | Downloads | Help & Docs | Blog
Login/Register
Search Human...
Human (GRCh38.p13)
Location: 8:66,868,488-67,868,488
Gene: CSPP1
Transcript: CSPP1-203
Transcript-based displays
Summary
Sequence
Exons
cDNA
Protein
Protein Information
Protein summary
Domains & features
Variants
3D Protein model
Genetic Variation
Variant table
Variant image
Haplotypes
Population comparison
Comparison image
External References
General identifiers
Oligo probes
Supporting evidence
ID History
Transcript history
Protein history
Configure this page
Custom tracks
Export data
Share this page
Bookmark this page
Transcript: CSPP1-203 ENST00000519668.1
Description
centrosome and spindle pole associated protein 1 [Source:HGNC Symbol;Acc:HGNC:26193]
Gene Synonyms
CSPP, FLJ22490, JBTS21
Location
Chromosome 8: 67,083,848-67,195,611 forward strand.
About this transcript
This transcript has 26 exons, is annotated with 23 domains and features, is associated with 26505 variant alleles and maps to 692 oligo probes.
Gene
This transcript is a product of gene ENSG00000104218.14
Hide transcript table
Show/hide columns (1 hidden)
Filter
Name | Transcript ID | bp | Protein | Biotype | CCDS | UniProt | RefSeq Match | Flags
CSPP1-201 | ENST00000262210.9 | 4367 | 1221aa | Protein coding | CCDS43744 | Q1MSJ5 | - | TSL:1 GENCODE basic APPRIS P3
CSPP1-203 | ENST00000519668.1 | 3715 | 876aa | Protein coding | CCDS78344 | Q1MSJ5 | - | TSL:1 GENCODE basic APPRIS ALT2
CSPP1-207 | ENST00000521919.5 | 567 | 154aa | Protein coding | - | E5RI67 | - | CDS 3' incomplete TSL:4
CSPP1-202 | ENST00000519163.6 | 3982 | 404aa | Nonsense mediated decay | - | F2Z2M5 | - | TSL:2
CSPP1-204 | ENST00000519701.5 | 569 | 66aa | Nonsense mediated decay | - | E5RGA5 | - | TSL:4
CSPP1-205 | ENST00000521168.5 | 1457 | No protein | lncRNA | - | - | - | TSL:1
CSPP1-206 | ENST00000521324.1 | 725 | No protein | lncRNA | - | - | - | TSL:3
Protein summary
Protein domains for ENSP00000430092.1
ENSP00000430092.1
MobiDB lite
Low complexity (Seq)
Coiled-coils (Ncoils)
PANTHER
PTHR21616:SF2
Centrosome and spindle pole associated protein 1
All sequence SNPs/i...
Sequence variants (dbSNP and all other sources)

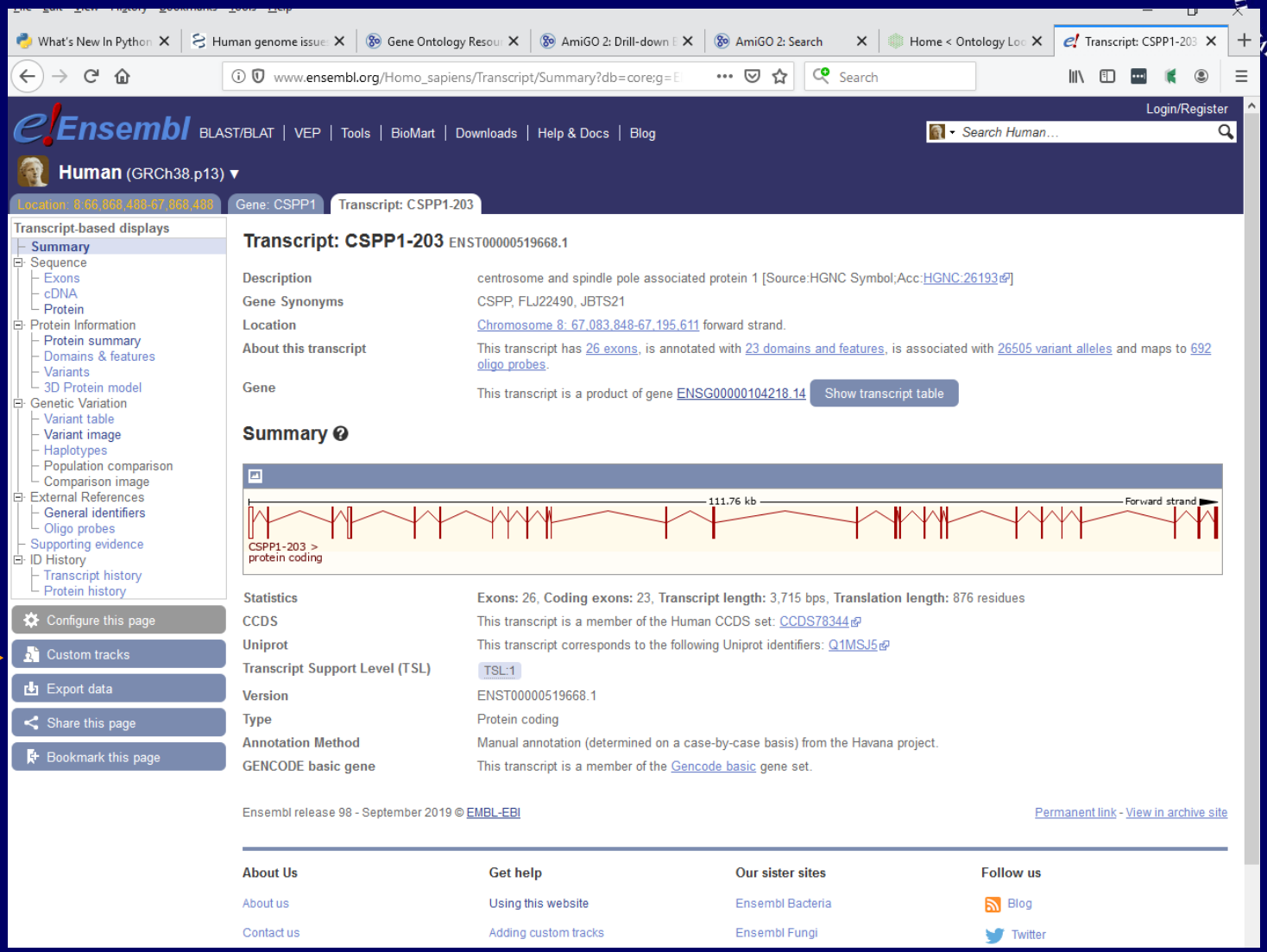

Transcript: CSPP1-203
www.ensembl.org/Homo_sapiens/Transcript/Variation_Transcript/Ima
Search
Genetic Variation
Variant table
Variant image
Haplotypes
Population comparison
Comparison image
External References
General identifiers
Oligo probes
Supporting evidence
ID History
Transcript history
Protein history
Configure this page
Custom tracks
Export data
Share this page
Bookmark this page
Variant image
222.17 kb
67.025Mb
67.050Mb
67.075Mb
67.100Mb
67.125Mb
67.150Mb
67.175Mb
67.200Mb
67.225Mb
Variant - COS...
Variant - COSMIC somatic mutations
All sequence S...
Sequence variants (dbSNP and all other sources)
Genes (Comprehensiv...
Genes (Comprehensiv...
ENST00000519668
CSPP1-203
Variant Legend
3 prime UTR variant
splice acceptor variant
stop gained
stop lost
inframe insertion
5 prime UTR variant
splice donor variant
frameshift variant
start lost
inframe deletion

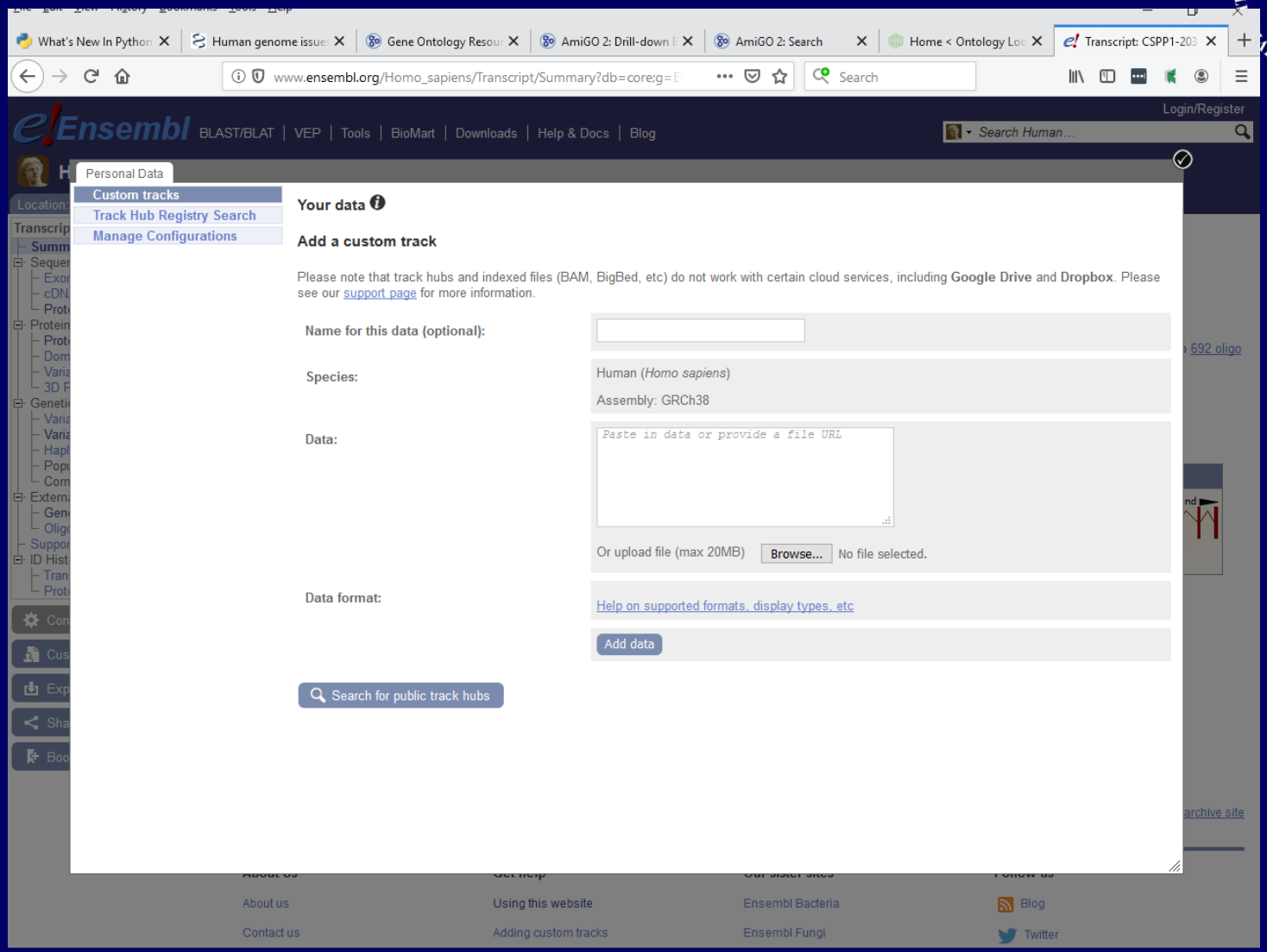

What's New In Python
Human genome issues
Gene Ontology Resou
AmiGO 2: Drill-down
AmiGO 2: Search
Home < Ontology Loc
Transcript: CSPP1-203
www.ensembl.org/Homo_sapiens/Transcript/Summary?db=core;g=E
Search
Ensembl
BLAST/BLAT | VEP | Tools | BioMart | Downloads | Help & Docs | Blog
Login/Register
Search Human...
Human (GRCh38.p13)
Location: 8:66,868,488-67,868,488
Gene: CSPP1
Transcript: CSPP1-203
Transcript-based displays
Summary
Sequence
Exons
cDNA
Protein
Protein Information
Protein summary
Domains & features
Variants
3D Protein model
Genetic Variation
Variant table
Variant image
Haplotypes
Population comparison
Comparison image
External References
General identifiers
Oligo probes
Supporting evidence
ID History
Transcript history
Protein history
Configure this page
Custom tracks
Export data
Share this page
Bookmark this page
Transcript: CSPP1-203 ENST00000519668.1
Description
centrosome and spindle pole associated protein 1 [Source:HGNC Symbol;Acc:HGNC:26193]
Gene Synonyms
CSPP, FLJ22490, JBTS21
Location
Chromosome 8: 67,083,848-67,195,611 forward strand.
About this transcript
This transcript has 26 exons, is annotated with 23 domains and features, is associated with 26505 variant alleles and maps to 692 oligo probes.
Gene
This transcript is a product of gene ENSG00000104218.14
Show transcript table
Summary
111.76 kb
Forward strand
CSPP1-203 >
protein coding
Statistics
Exons: 26, Coding exons: 23, Transcript length: 3,715 bps, Translation length: 876 residues
CCDS
This transcript is a member of the Human CCDS set: CCDS78344
Uniprot
This transcript corresponds to the following Uniprot identifiers: Q1MSJ5
Transcript Support Level (TSL)
TSL:1
Version
ENST00000519668.1
Type
Protein coding
Annotation Method
Manual annotation (determined on a case-by-case basis) from the Havana project.
GENCODE basic gene
This transcript is a member of the Gencode basic gene set.
Ensembl release 98 - September 2019 © EMBL-EBI
Permanent link - View in archive site
About Us
About us
Contact us
Get help
Using this website
Adding custom tracks
Our sister sites
Ensembl Bacteria
Ensembl Fungi
Follow us
Blog
Twitter


What's New In Python
Human genome issues
Gene Ontology Resou
AmiGO 2: Drill-down
AmiGO 2: Search
Home < Ontology Loc
Transcript: CSPP1-203
www.ensembl.org/Homo_sapiens/Transcript/Summary?db=core;g=E
Search
Ensembl
BLAST/BLAT | VEP | Tools | BioMart | Downloads | Help & Docs | Blog
Login/Register
Search Human...
Export data
Export Configuration - Feature List
Tip
For sequence export, please go to the relevant sequence page (see lefthand menu) and use the new "Download sequence" button
Transcript to export:
ENST00000519668.1 (CSPP1-203)
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:
692 oligo
archive site
About us
Contact us
Using this website
Adding custom tracks
Ensembl Bacteria
Ensembl Fungi
Blog
Twitter

Az osztott annotációs rendszer (DAS)

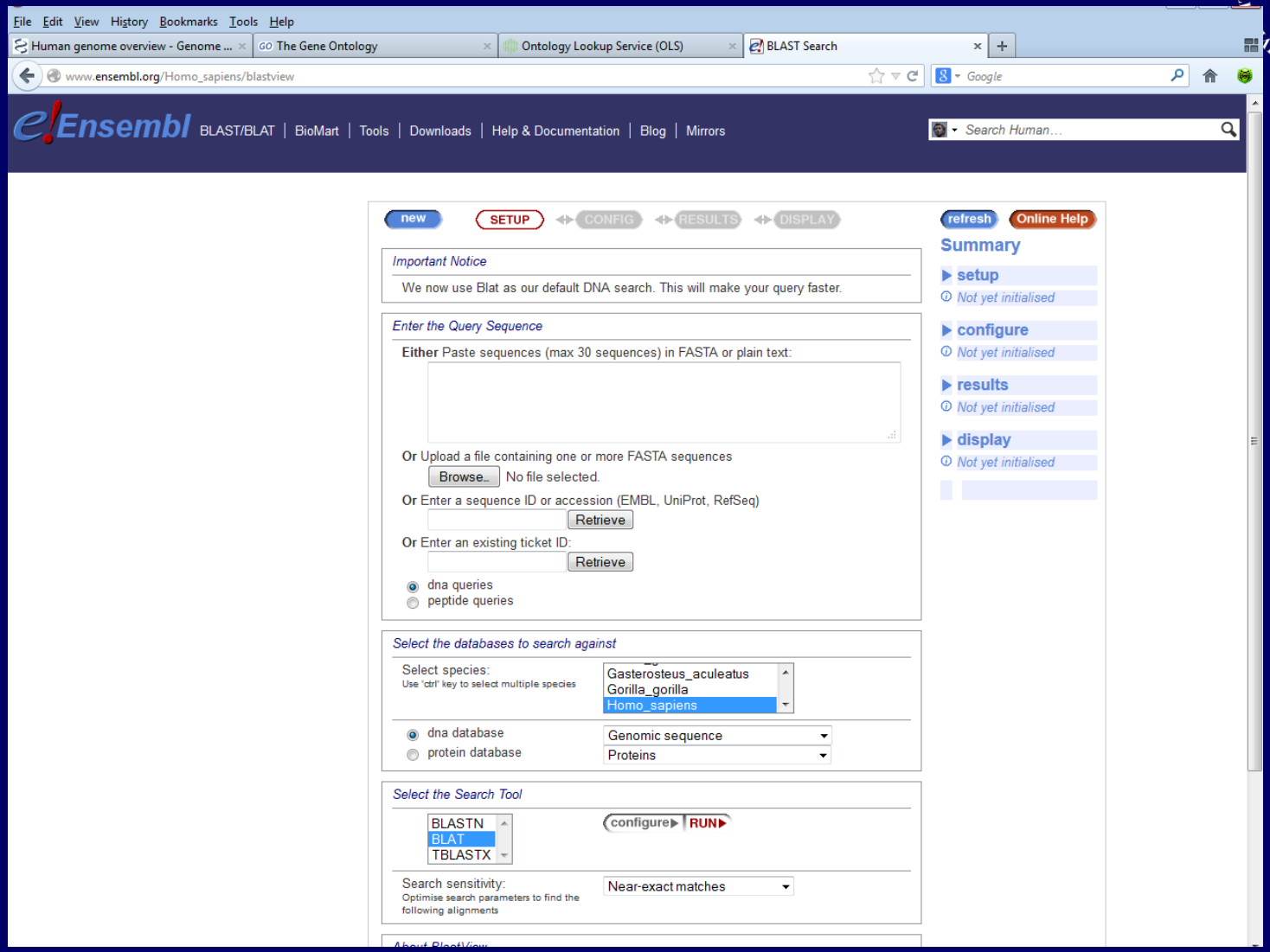
- Célja:
  - Saját adatok feltüntetése az ENSEMBL felületen
  - Adatok megosztása együttműködők közt
- Feltöltéshez használható egy külön fül a felületen
- A formátum kötött, alapvetően szöveges (ld. „Help”)
- ENSEMBL regisztráció és belépés segít

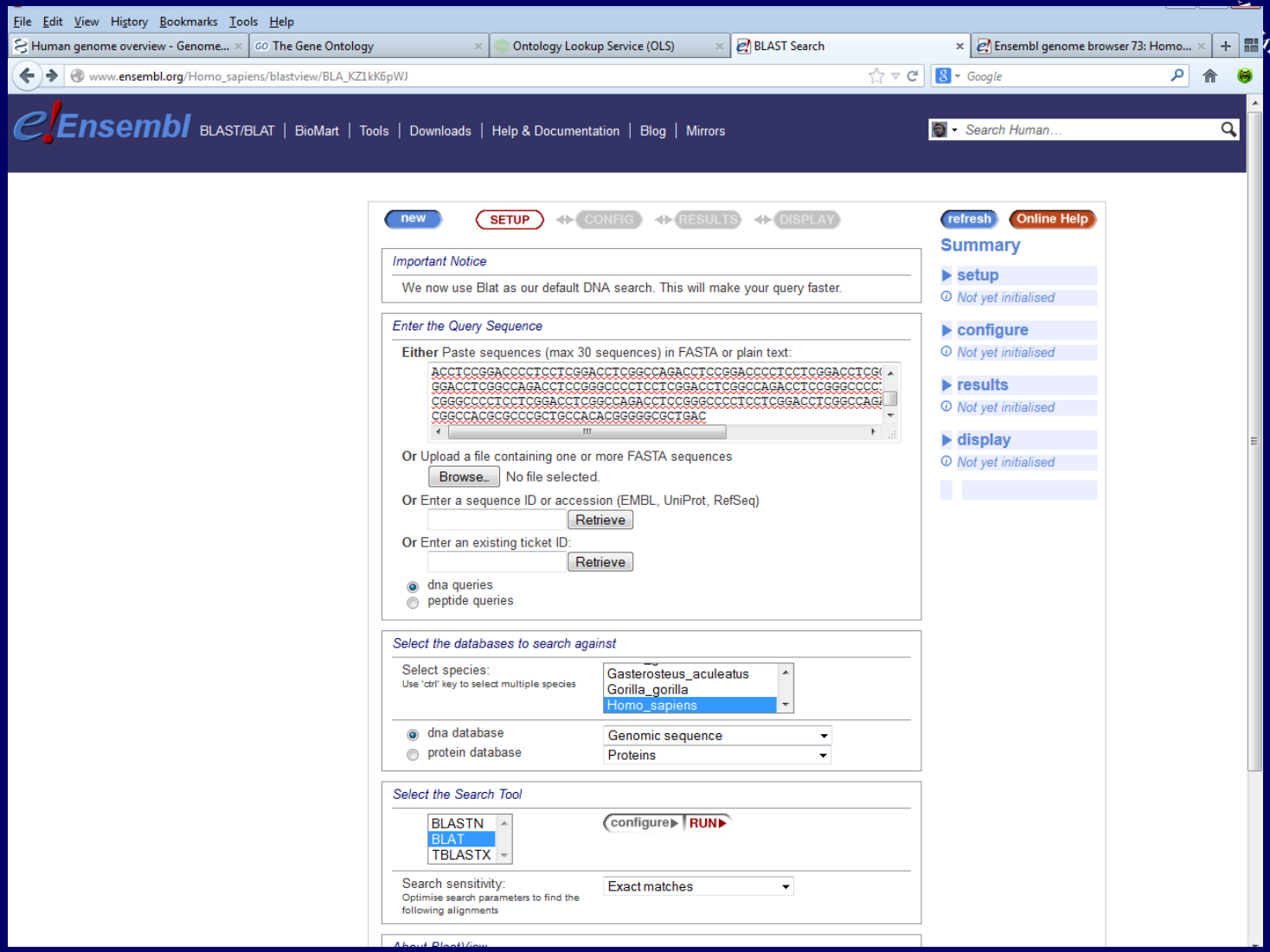

What's New In Python
Human genome issues
Gene Ontology Resou
AmiGO 2: Drill-down
AmiGO 2: Search
Home < Ontology Loc
Transcript: CSPP1-203
www.ensembl.org/Homo_sapiens/Transcript/Summary?db=core;g=E
Search
Ensembl
BLAST/BLAT | VEP | Tools | BioMart | Downloads | Help & Docs | Blog
Login/Register
Search Human...
Human (GRCh38.p13)
Location: 8:66,868,488-67,868,488
Gene: CSPP1
Transcript: CSPP1-203
Transcript-based displays
Summary
Sequence
Exons
cDNA
Protein
Protein Information
Protein summary
Domains & features
Variants
3D Protein model
Genetic Variation
Variant table
Variant image
Haplotypes
Population comparison
Comparison image
External References
General identifiers
Oligo probes
Supporting evidence
ID History
Transcript history
Protein history
Configure this page
Custom tracks
Export data
Share this page
Bookmark this page
Transcript: CSPP1-203 ENST00000519668.1
Description
centrosome and spindle pole associated protein 1 [Source:HGNC Symbol;Acc:HGNC:26193]
Gene Synonyms
CSPP, FLJ22490, JBTS21
Location
Chromosome 8: 67,083,848-67,195,611 forward strand.
About this transcript
This transcript has 26 exons, is annotated with 23 domains and features, is associated with 26505 variant alleles and maps to 692 oligo probes.
Gene
This transcript is a product of gene ENSG00000104218.14
Show transcript table
Summary
111.76 kb
Forward strand
CSPP1-203 >
protein coding
Statistics
Exons: 26, Coding exons: 23, Transcript length: 3,715 bps, Translation length: 876 residues
CCDS
This transcript is a member of the Human CCDS set: CCDS78344
Uniprot
This transcript corresponds to the following Uniprot identifiers: Q1MSJ5
Transcript Support Level (TSL)
TSL:1
Version
ENST00000519668.1
Type
Protein coding
Annotation Method
Manual annotation (determined on a case-by-case basis) from the Havana project.
GENCODE basic gene
This transcript is a member of the Gencode basic gene set.
Ensembl release 98 - September 2019 © EMBL-EBI
Permanent link - View in archive site
About Us
About us
Contact us
Get help
Using this website
Adding custom tracks
Our sister sites
Ensembl Bacteria
Ensembl Fungi
Follow us
Blog
Twitter

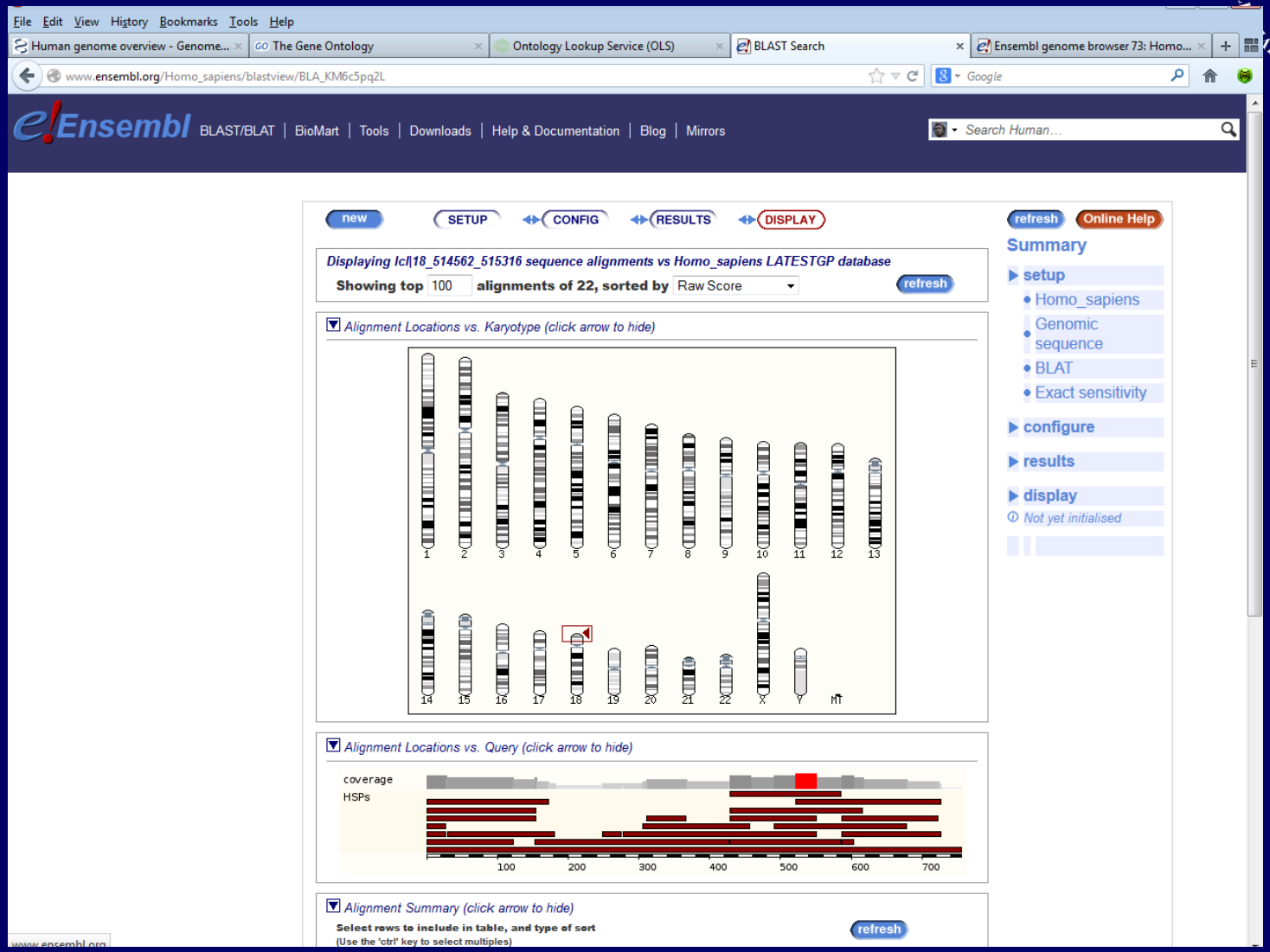

What's New In Python
Human genome issue
Gene Ontology Resou
AmiGO 2: Drill-down
AmiGO 2: Search
Home < Ontology Loc
Transcript: CSPP1-203
www.ensembl.org/Homo_sapiens/Transcript/Summary?db=core;g=E
Search
Ensembl
BLAST/BLAT | VEP | Tools | BioMart | Downloads | Help & Docs | Blog
Login/Register
Search Human...
Personal Data
Custom tracks
Track Hub Registry Search
Manage Configurations
Your data
Add a custom track
Please note that track hubs and indexed files (BAM, BigBed, etc) do not work with certain cloud services, including Google Drive and Dropbox. Please see our support page for more information.
Name for this data (optional):
Species:
Human (Homo sapiens)
Assembly: GRCh38
Data:
Paste in data or provide a file URL
Or upload file (max 20MB)
Browse...
No file selected.
Data format:
Help on supported formats, display types, etc
Add data
Search for public track hubs
692 oligo
archive site
About us
Contact us
Using this website
Adding custom tracks
Ensembl Bacteria
Ensembl Fungi
Blog
Twitter

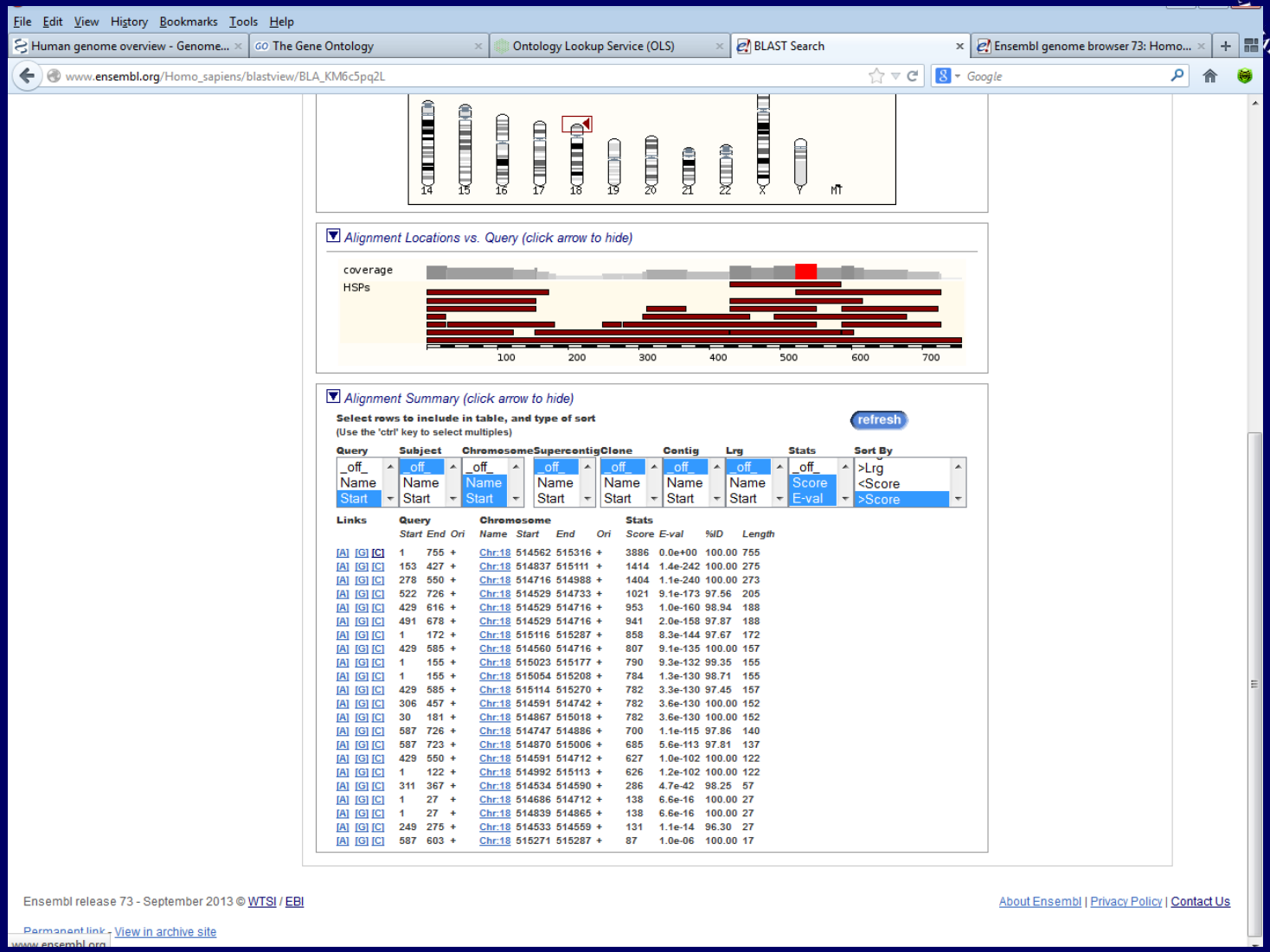
Kapcsolódó eszközök

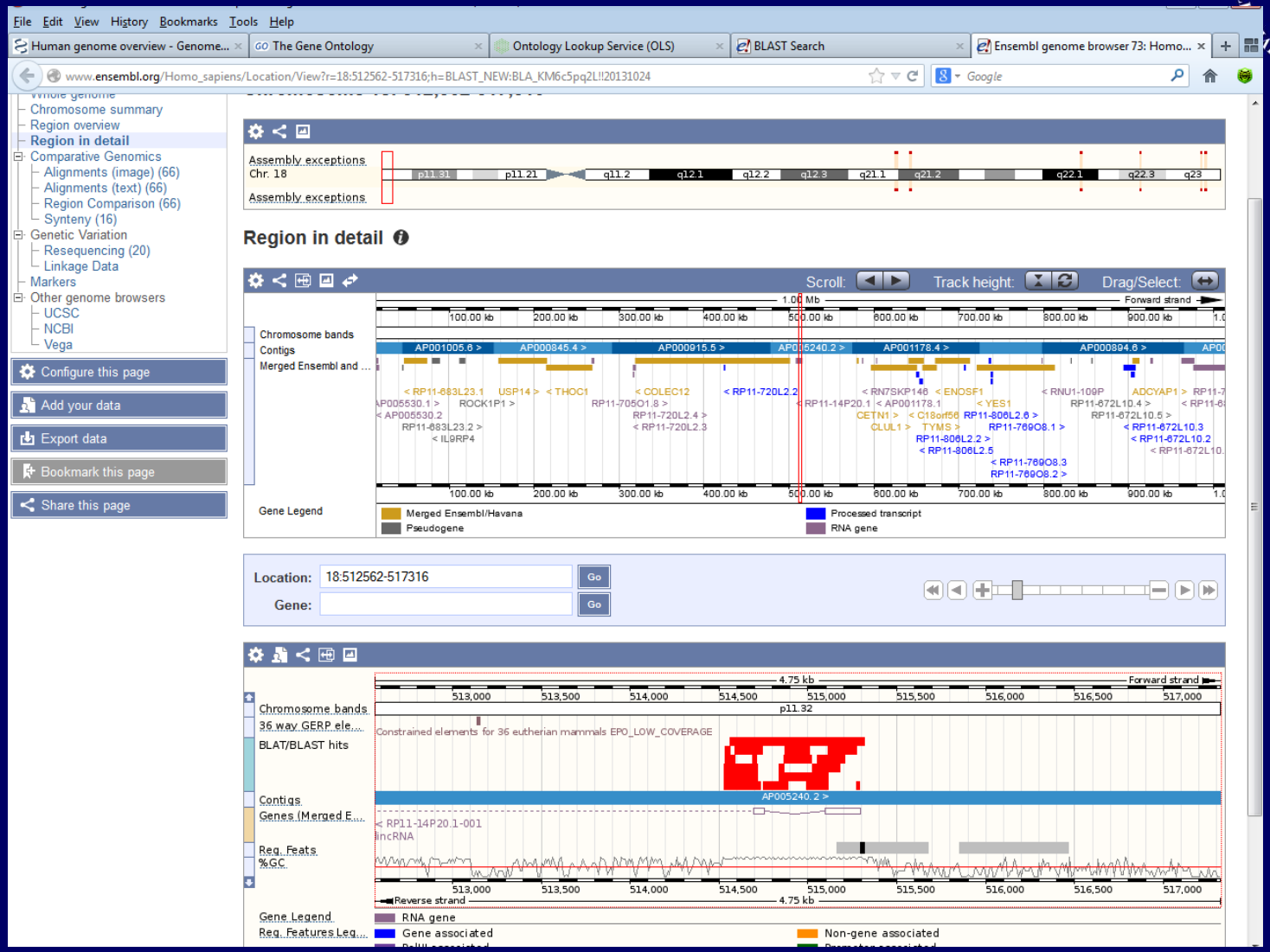
- BLAST/BLAT
- BioMart
- Adatletöltési lehetőségek


File Edit View History Bookmarks Tools Help
Human genome overview - Genome ...
The Gene Ontology
Ontology Lookup Service (OLS)
BLAST Search
www.ensembl.org/Homo_sapiens/blastview
Google
e!Ensembl
BLAST/BLAT | BioMart | Tools | Downloads | Help & Documentation | Blog | Mirrors
Search Human...
new
SETUP
CONFIG
RESULTS
DISPLAY
refresh
Online Help
Summary
setup
Not yet initialised
configure
Not yet initialised
results
Not yet initialised
display
Not yet initialised
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
Browse...
No file selected.
Or Enter a sequence ID or accession (EMBL, UniProt, RefSeq)
Retrieve
Or Enter an existing ticket ID:
Retrieve
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


File Edit View History Bookmarks Tools Help
Human genome overview - Genome...
The Gene Ontology
Ontology Lookup Service (OLS)
BLAST Search
Ensembl genome browser 73: Homo...
www.ensembl.org/Homo_sapiens/blastview/BLA_KZ1kK6pWJ
Ensembl BLAST/BLAT | BioMart | Tools | Downloads | Help & Documentation | Blog | Mirrors
Search Human...
new
SETUP
CONFIG
RESULTS
DISPLAY
refresh
Online Help
Summary
setup
Not yet initialised
configure
Not yet initialised
results
Not yet initialised
display
Not yet initialised
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:
ACCTCCGGACCCCTCCTCGGACCTCGGCCAGACCTCCGGACCCCTCCTCGGACCTCG
GGACCTCGGCCAGACCTCCGGGCCCCTCCTCGGACCTCGGCCAGACCTCCGGGCCCC
CGGGCCCCTCCTCGGACCTCGGCCAGACCTCCGGGCCCCTCCTCGGACCTCGGCCAG
CGGCCACGCGCCCGCTGCCACACGGGGGCGCTGAC
Or Upload a file containing one or more FASTA sequences
Browse...
No file selected.
Or Enter a sequence ID or accession (EMBL, UniProt, RefSeq)
Retrieve
Or Enter an existing ticket ID:
Retrieve
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
Exact matches

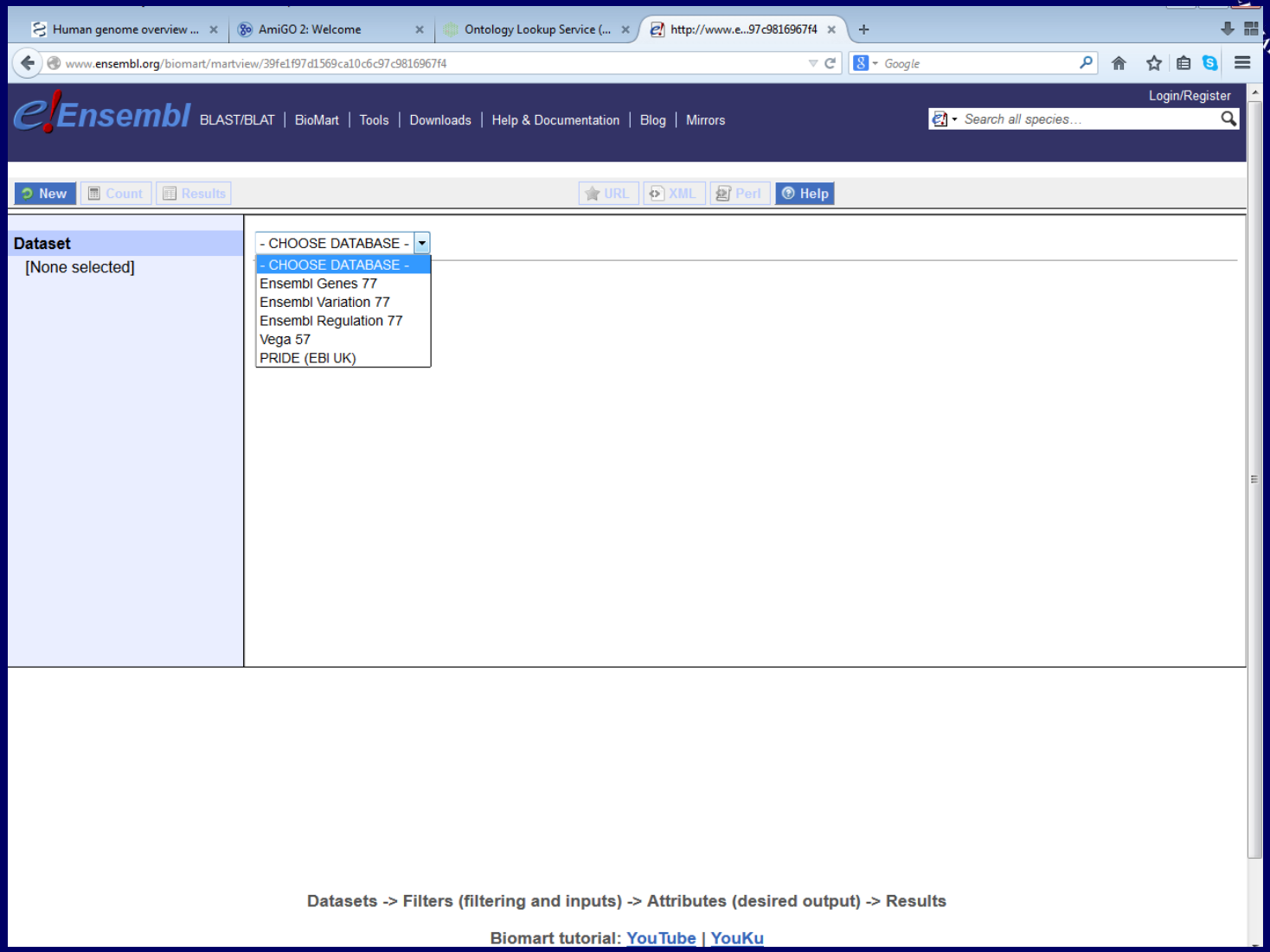

File Edit View History Bookmarks Tools Help
Human genome overview - Genome...
The Gene Ontology
Ontology Lookup Service (OLS)
BLAST Search
Ensembl genome browser 73: Homo...
www.ensembl.org/Homo_sapiens/blastview/BLA_KM6c5pq2L
Ensembl BLAST/BLAT | BioMart | Tools | Downloads | Help & Documentation | 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 lcl|18_514562_515316 sequence alignments vs Homo_sapiens LATESTGP database
Showing top 100 alignments of 22, sorted by Raw Score
refresh
Alignment Locations vs. Karyotype (click arrow to hide)
1 2 3 4 5 6 7 8 9 10 11 12 13
14 15 16 17 18 19 20 21 22 X Y MT
Alignment Locations vs. Query (click arrow to hide)
coverage
HSPs
100 200 300 400 500 600 700
Alignment Summary (click arrow to hide)
Select rows to include in table, and type of sort
(Use the 'ctrl' key to select multiples)
refresh

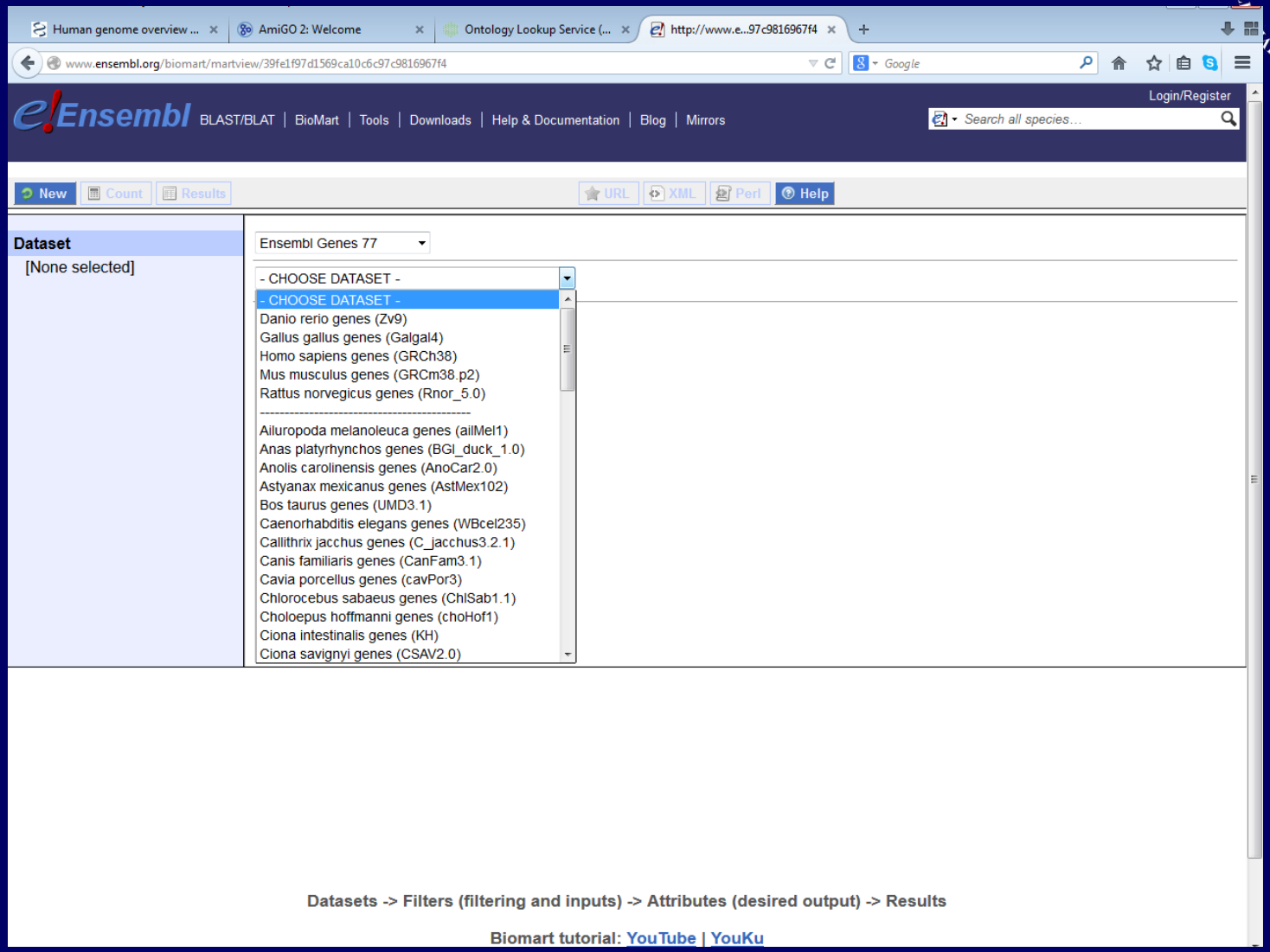

File Edit View History Bookmarks Tools Help
Human genome overview - Genome...
The Gene Ontology
Ontology Lookup Service (OLS)
BLAST Search
Ensembl genome browser 73: Homo...
www.ensembl.org/Homo_sapiens/blastview/BLA_KM6c5pq2L
14 15 16 17 18 19 20 21 22 X Y MT
Alignment Locations vs. Query (click arrow to hide)
coverage
HSPs
100 200 300 400 500 600 700
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
Links Query Chromosome 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

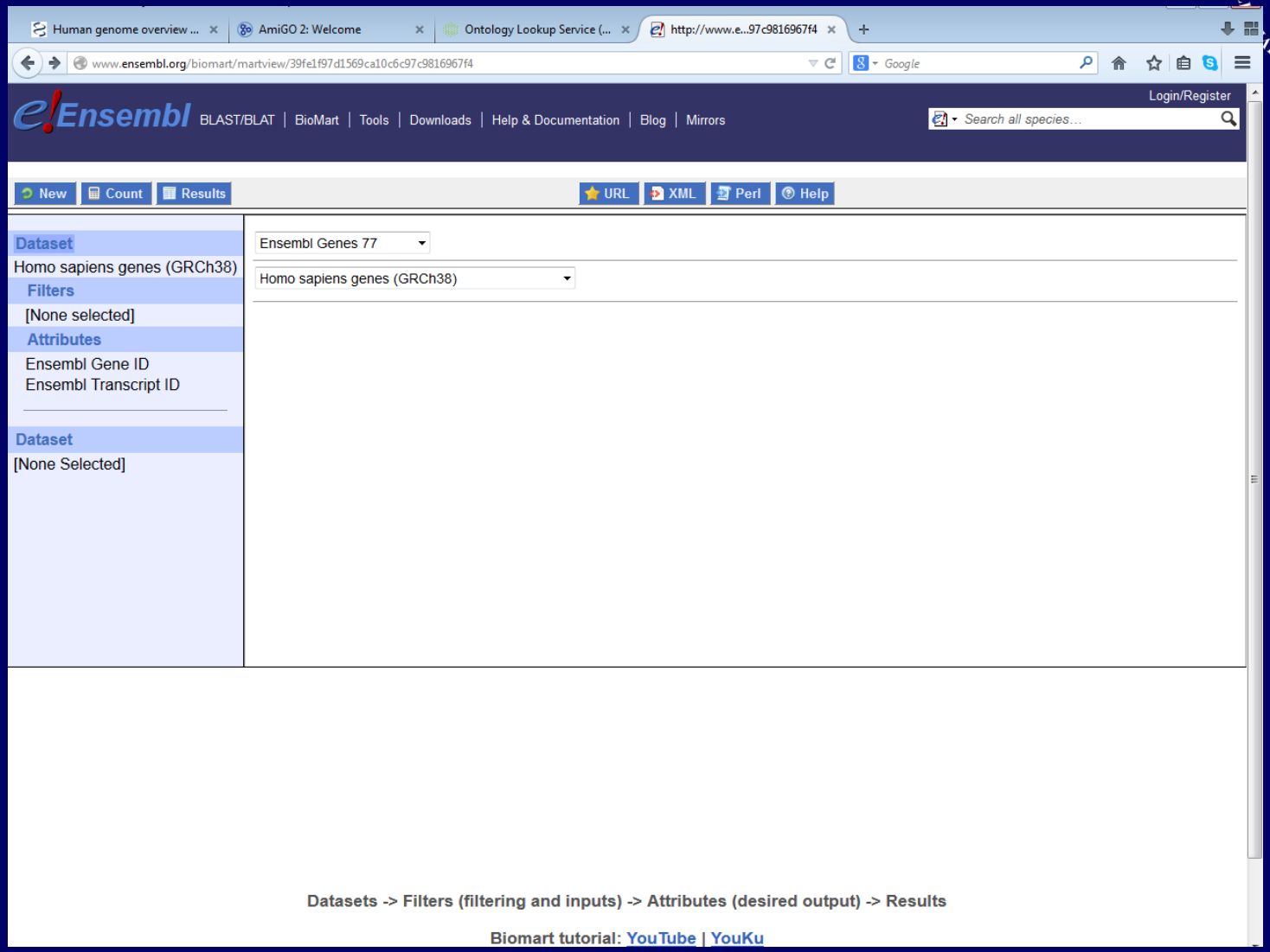

File Edit View History Bookmarks Tools Help
Human genome overview - Genome...
The Gene Ontology
Ontology Lookup Service (OLS)
BLAST Search
Ensembl genome browser 73: Homo...
www.ensembl.org/Homo_sapiens/Location/View?r=18:512562-517316;h=BLAST_NEW:BLA_KM6c5pq2L!!20131024
Chromosome summary
Region overview
Region in detail
Comparative Genomics
Alignments (image) (66)
Alignments (text) (66)
Region Comparison (66)
Synteny (16)
Genetic Variation
Resequencing (20)
Linkage Data
Markers
Other genome browsers
UCSC
NCBI
Vega
Configure this page
Add your data
Export data
Bookmark this page
Share this page
Assembly exceptions
Chr. 18
Region in detail
Scroll:
Track height:
Drag/Select:
Chromosome bands
Contigs
Merged Ensembl and ...
Gene Legend
Merged Ensembl/Havana
Pseudogene
Processed transcript
RNA gene
Location: 18:512562-517316
Gene:
Go
Chromosome bands
36 way GERP ele...
BLAT/BLAST hits
Constrained elements for 36 eutherian mammals EPO_LOW_COVERAGE
Contigs
Genes (Merged E...
Reg. Feats
%GC
Gene Legend
Reg. Features Leg...
RNA gene
Gene associated
Non-gene associated

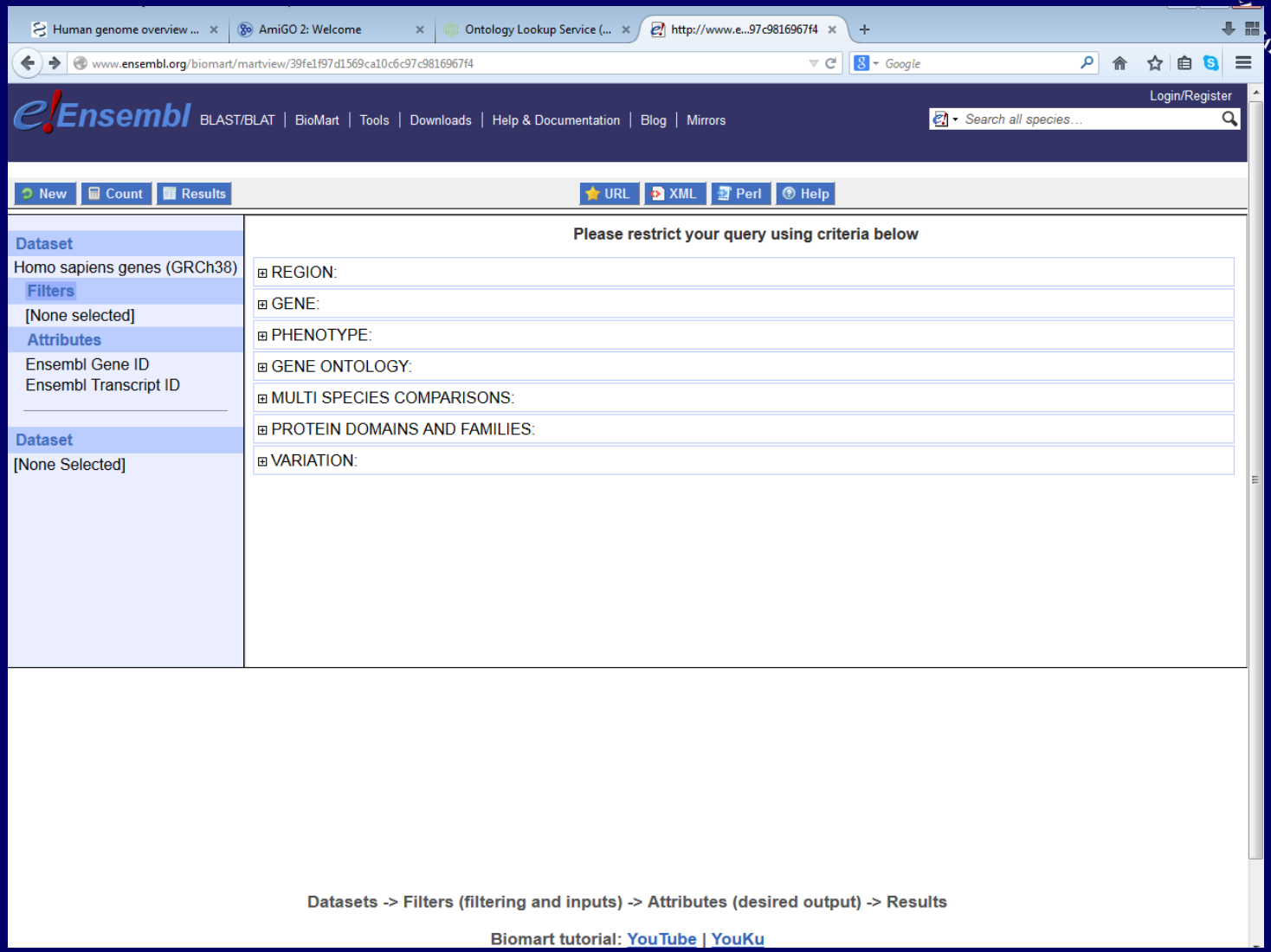
Tömeges lekérdezés

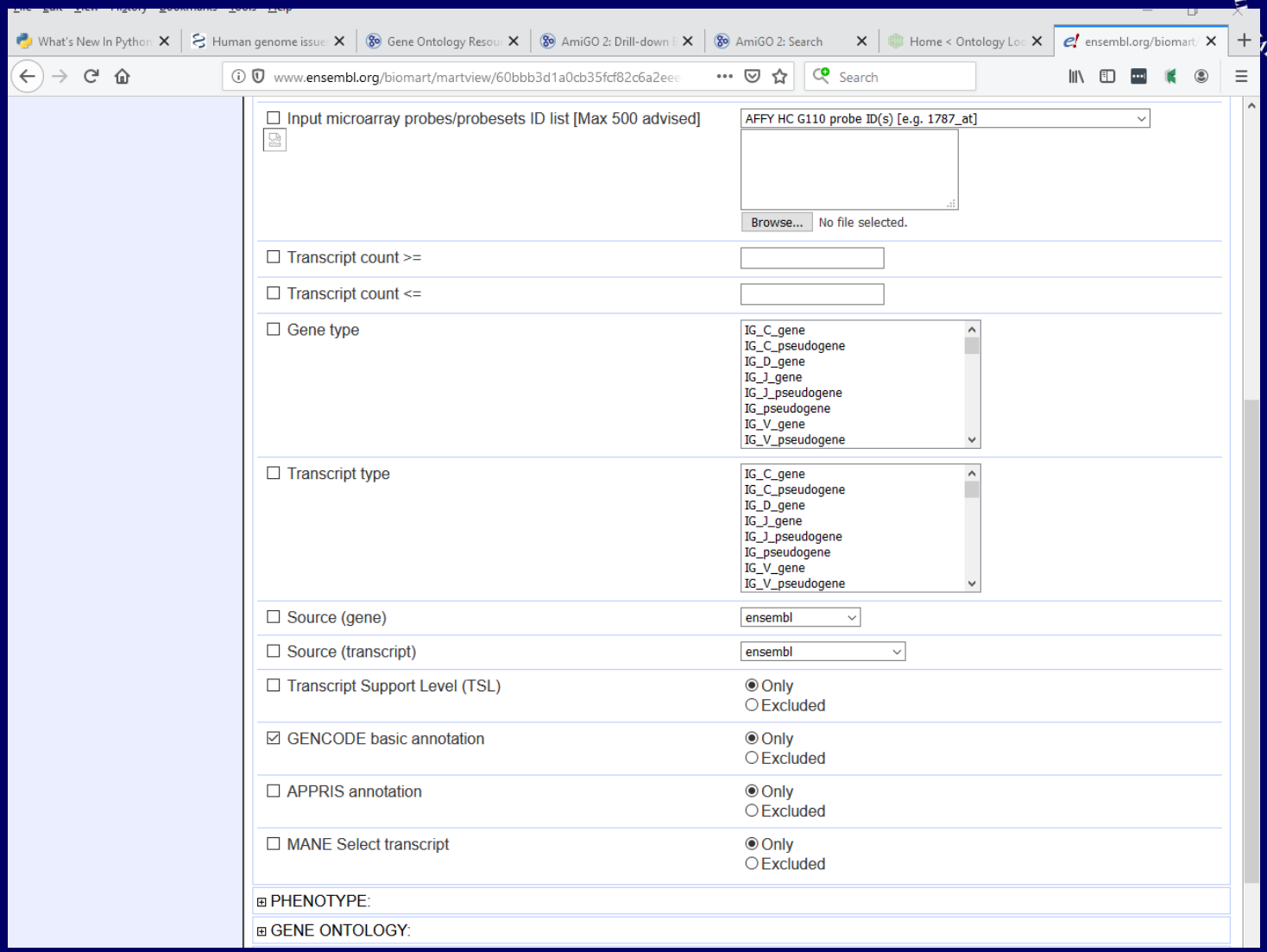
- BioMart rendszer
- Web-felületen át elérhető
- Az adattartalom azonos az interaktív felületem megjelenítettel
- Szöveges eredményt ad
- Nagy adattömeget fogunk kapni eredményül

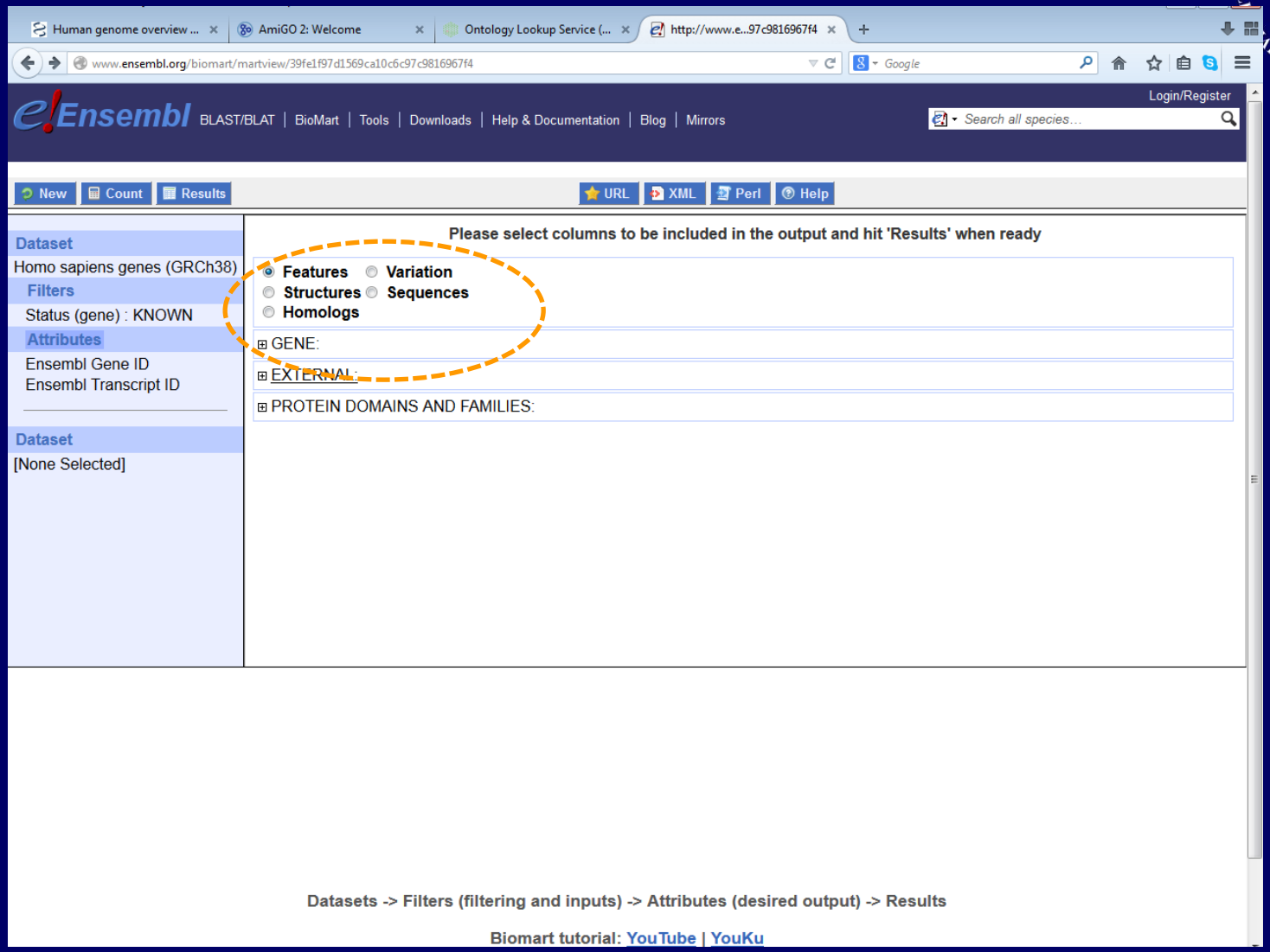

Human genome overview ...
AmiGO 2: Welcome
Ontology Lookup Service (...
http://www.e...97c9816967f4
www.ensembl.org/biomart/martview/39fe1f97d1569ca10c6c97c9816967f4
Google
e!Ensembl
BLAST/BLAT | BioMart | Tools | Downloads | Help & Documentation | Blog | Mirrors
Login/Register
Search all species...
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

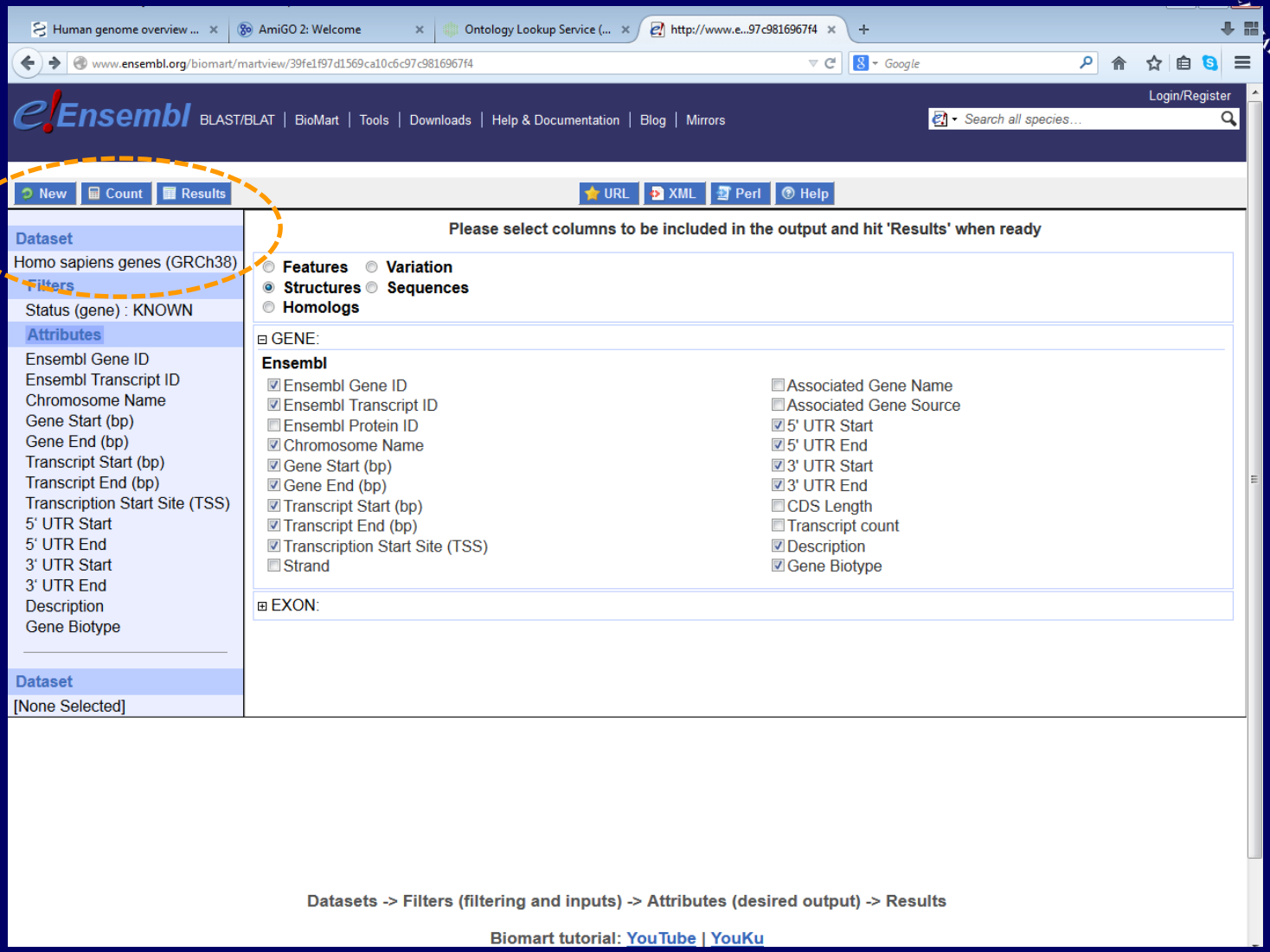

Human genome overview ...
AmiGO 2: Welcome
Ontology Lookup Service (...
http://www.e...97c9816967f4
www.ensembl.org/biomart/martview/39fe1f97d1569ca10c6c97c9816967f4
Ensembl
BLAST/BLAT | BioMart | Tools | Downloads | Help & Documentation | Blog | Mirrors
Login/Register
Search all species...
New
Count
Results
URL
XML
Perl
Help
Dataset
[None selected]
Ensembl Genes 77
- CHOOSE DATASET -
- CHOOSE DATASET -
Danio rerio genes (Zv9)
Gallus gallus genes (Galgal4)
Homo sapiens genes (GRCh38)
Mus musculus genes (GRCm38.p2)
Rattus norvegicus genes (Rnor_5.0)
Ailuropoda melanoleuca genes (ailMel1)
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

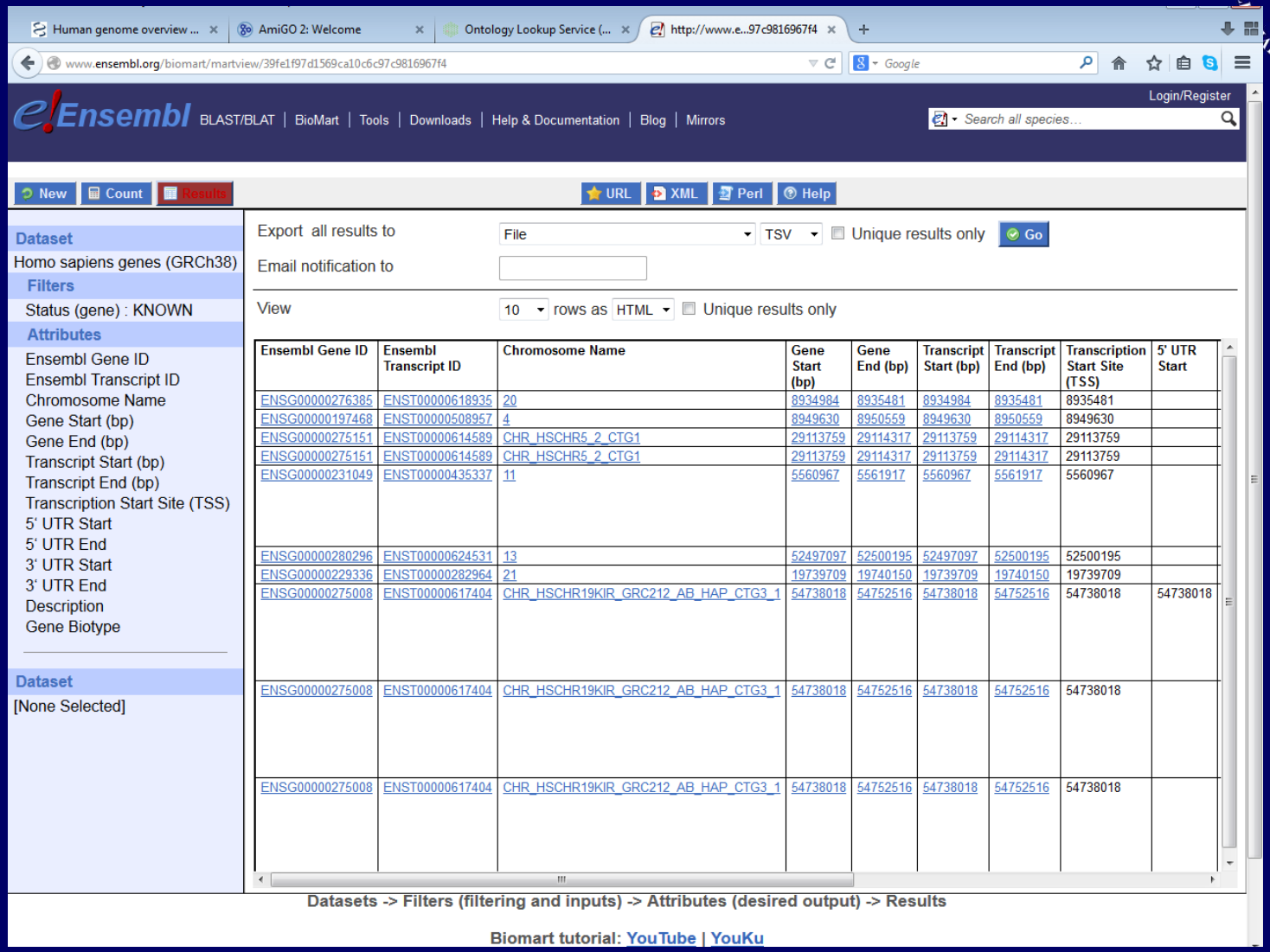

Human genome overview ...
AmiGO 2: Welcome
Ontology Lookup Service (...
http://www.e...97c9816967f4
www.ensembl.org/biomart/martview/39fe1f97d1569ca10c6c97c9816967f4
Ensembl
BLAST/BLAT | BioMart | Tools | Downloads | Help & Documentation | Blog | Mirrors
Login/Register
Search all species...
New
Count
Results
URL
XML
Perl
Help
Dataset
Homo sapiens genes (GRCh38)
Filters
[None selected]
Attributes
Ensembl Gene ID
Ensembl Transcript ID
Dataset
[None Selected]
Ensembl Genes 77
Homo sapiens genes (GRCh38)
Datasets -> Filters (filtering and inputs) -> Attributes (desired output) -> Results
Biomart tutorial: YouTube | YouKu

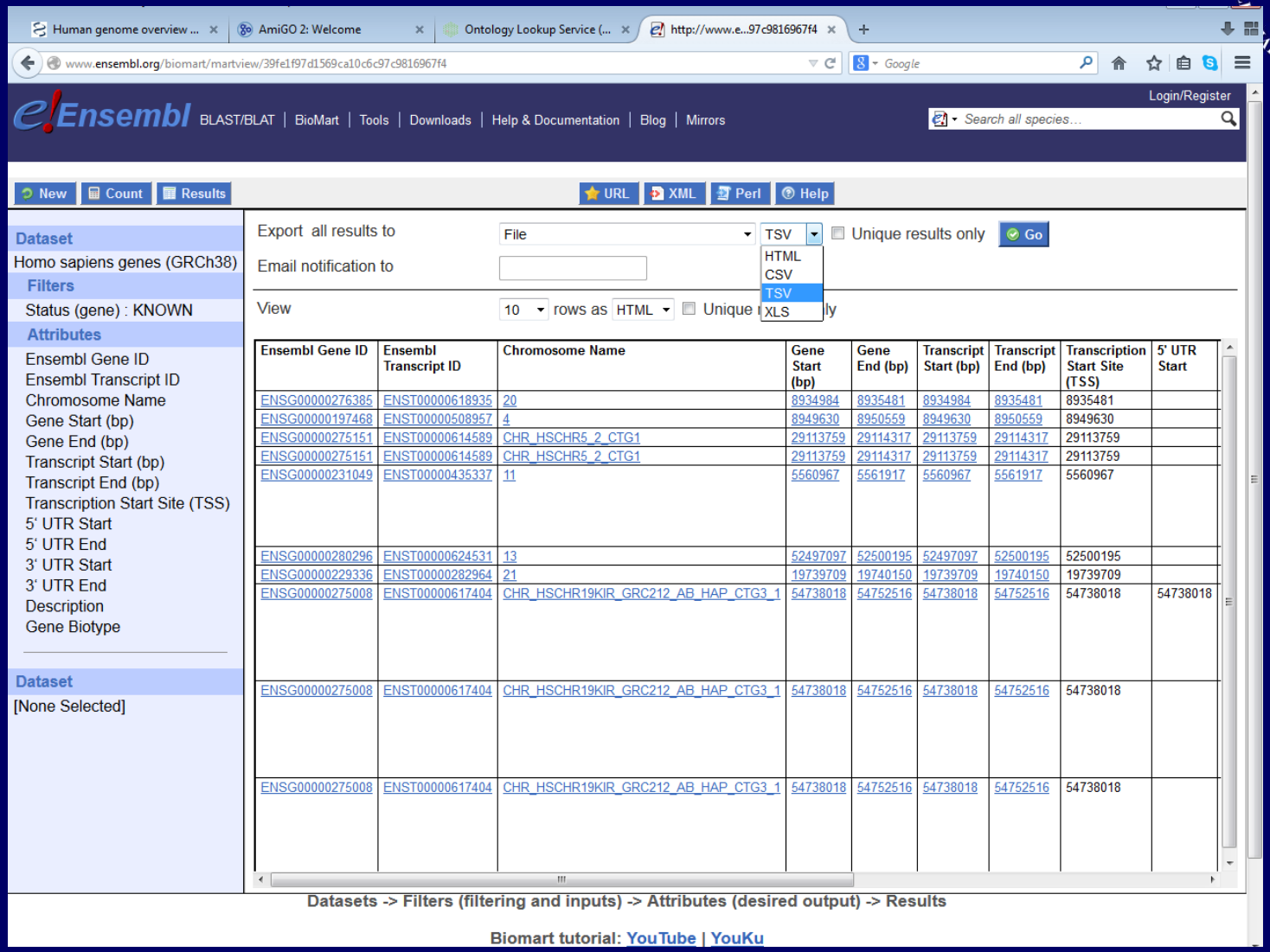

Human genome overview ...
AmiGO 2: Welcome
Ontology Lookup Service (...
http://www.e...97c9816967f4
www.ensembl.org/biomart/martview/39fe1f97d1569ca10c6c97c9816967f4
Ensembl
BLAST/BLAT | BioMart | Tools | Downloads | Help & Documentation | Blog | Mirrors
Login/Register
Search all species...
New
Count
Results
URL
XML
Perl
Help
Dataset
Homo sapiens genes (GRCh38)
Filters
[None selected]
Attributes
Ensembl Gene ID
Ensembl Transcript ID
Dataset
[None Selected]
Please restrict your query using criteria below
REGION:
GENE:
PHENOTYPE:
GENE ONTOLOGY:
MULTI SPECIES COMPARISONS:
PROTEIN DOMAINS AND FAMILIES:
VARIATION:
Datasets -> Filters (filtering and inputs) -> Attributes (desired output) -> Results
Biomart tutorial: YouTube | YouKu

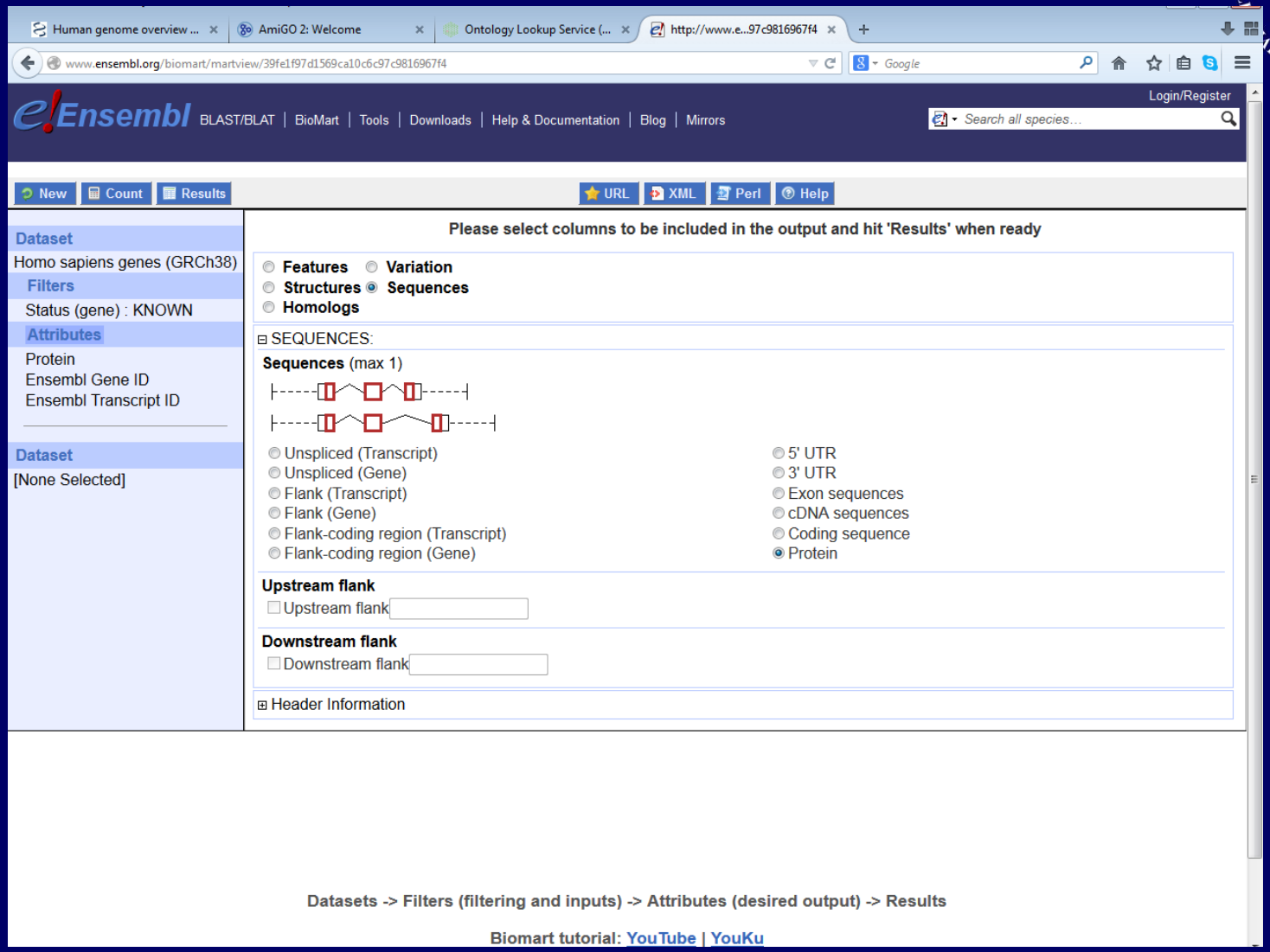

What's New In Python
Human genome issue
Gene Ontology Resou
AmiGO 2: Drill-down
AmiGO 2: Search
Home < Ontology Loc
ensembl.org/biomart
www.ensembl.org/biomart/martview/60bbb3d1a0cb35fcf82c6a2ee
Search
Input microarray probes/probesets ID list [Max 500 advised]
AFFY HC G110 probe ID(s) [e.g. 1787_at]
Browse...
No file selected.
Transcript count >=
Transcript count <=
Gene type
IG_C_gene
IG_C_pseudogene
IG_D_gene
IG_J_gene
IG_J_pseudogene
IG_pseudogene
IG_V_gene
IG_V_pseudogene
Transcript type
IG_C_gene
IG_C_pseudogene
IG_D_gene
IG_J_gene
IG_J_pseudogene
IG_pseudogene
IG_V_gene
IG_V_pseudogene
Source (gene)
ensembl
Source (transcript)
ensembl
Transcript Support Level (TSL)
Only
Excluded
GENCODE basic annotation
Only
Excluded
APPRIS annotation
Only
Excluded
MANE Select transcript
Only
Excluded
PHENOTYPE:
GENE ONTOLOGY:

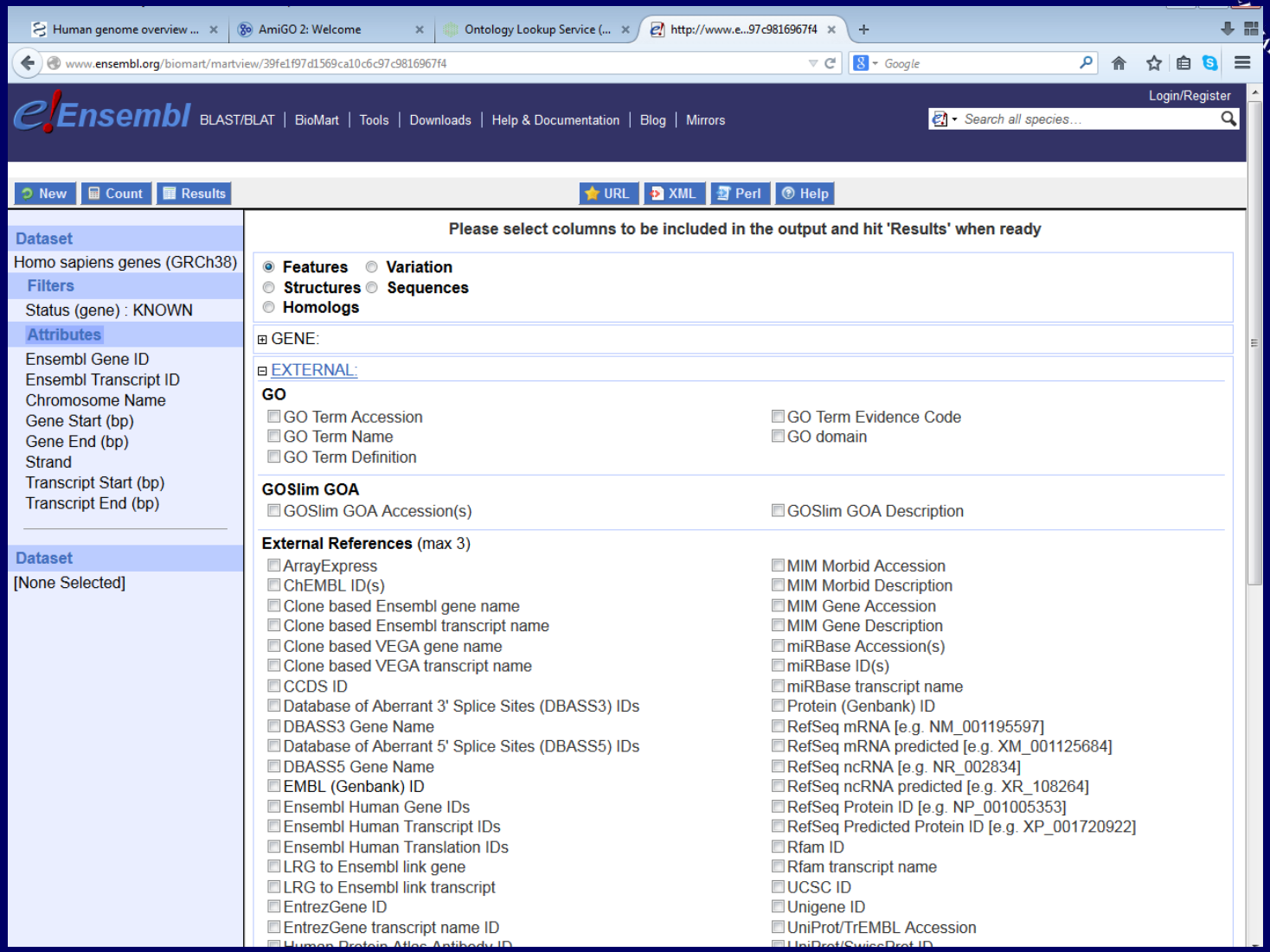

Human genome overview ...
AmiGO 2: Welcome
Ontology Lookup Service (...
http://www.e...97c9816967f4
www.ensembl.org/biomart/martview/39fe1f97d1569ca10c6c97c9816967f4
Ensembl
BLAST/BLAT | BioMart | Tools | Downloads | Help & Documentation | Blog | Mirrors
Login/Register
Search all species...
New
Count
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:
EXTERNAL:
PROTEIN DOMAINS AND FAMILIES:
Datasets -> Filters (filtering and inputs) -> Attributes (desired output) -> Results
Biomart tutorial: YouTube | YouKu

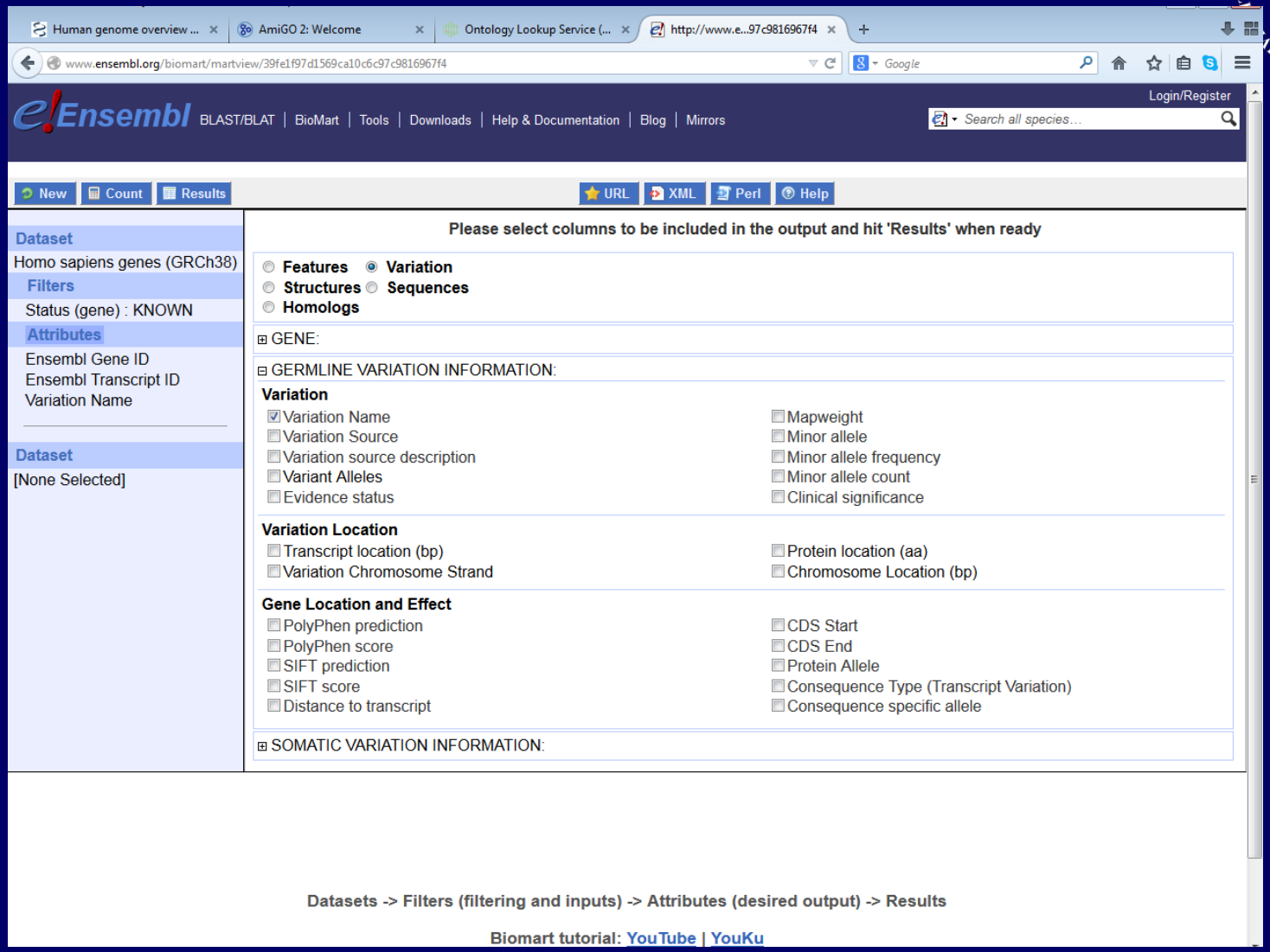

Human genome overview ...
AmiGO 2: Welcome
Ontology Lookup Service (...
http://www.e...97c9816967f4
www.ensembl.org/biomart/martview/39fe1f97d1569ca10c6c97c9816967f4
Ensembl
BLAST/BLAT | BioMart | Tools | Downloads | Help & Documentation | Blog | Mirrors
Login/Register
Search all species...
New
Count
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)
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]
Please select columns to be included in the output and hit 'Results' when ready
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

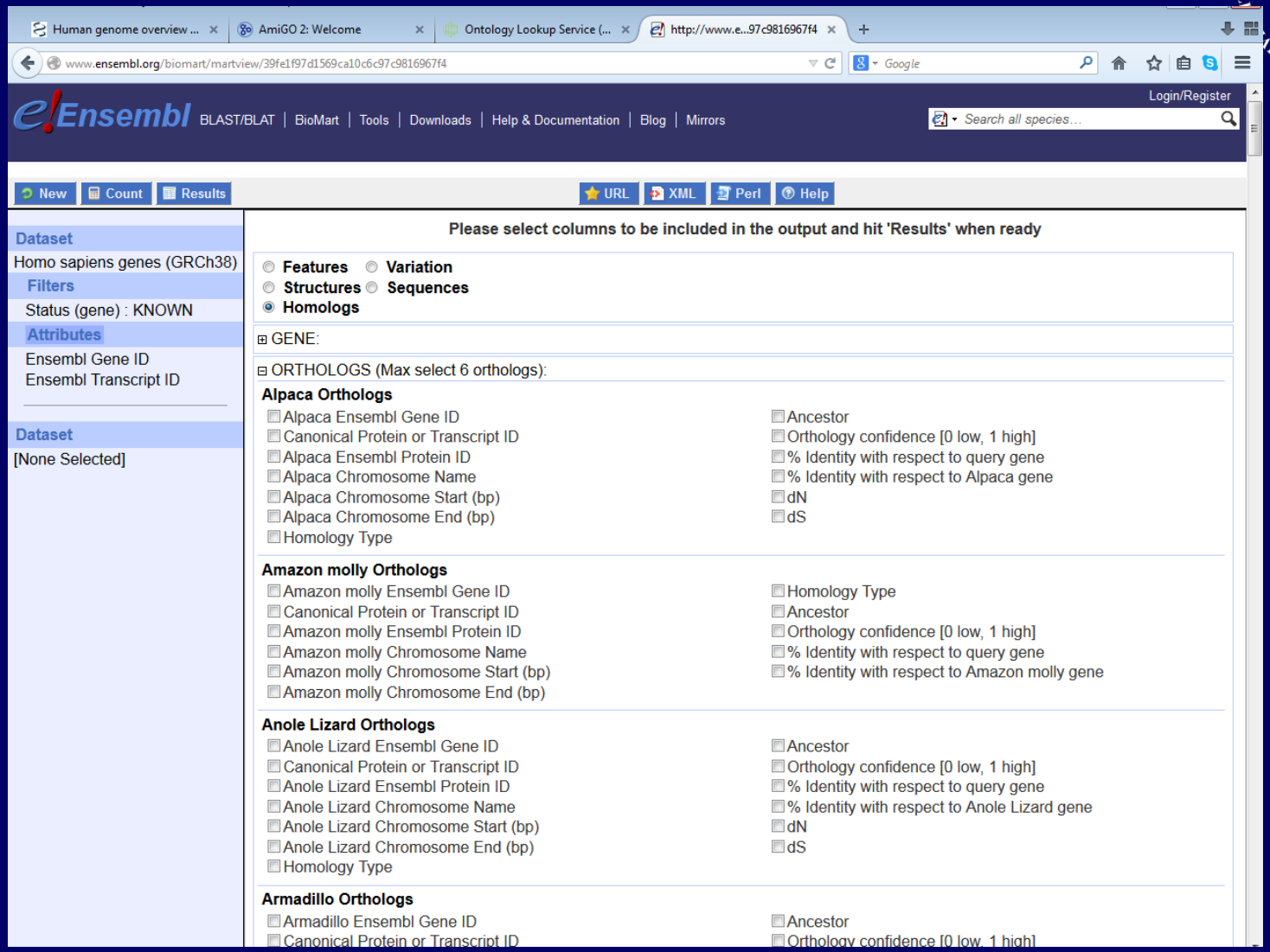

Human genome overview ...
AmiGO 2: Welcome
Ontology Lookup Service (...
http://www.e...97c9816967f4
www.ensembl.org/biomart/martview/39fe1f97d1569ca10c6c97c9816967f4
Ensembl
BLAST/BLAT | BioMart | Tools | Downloads | Help & Documentation | Blog | Mirrors
Login/Register
Search all species...
New
Count
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)
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]
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

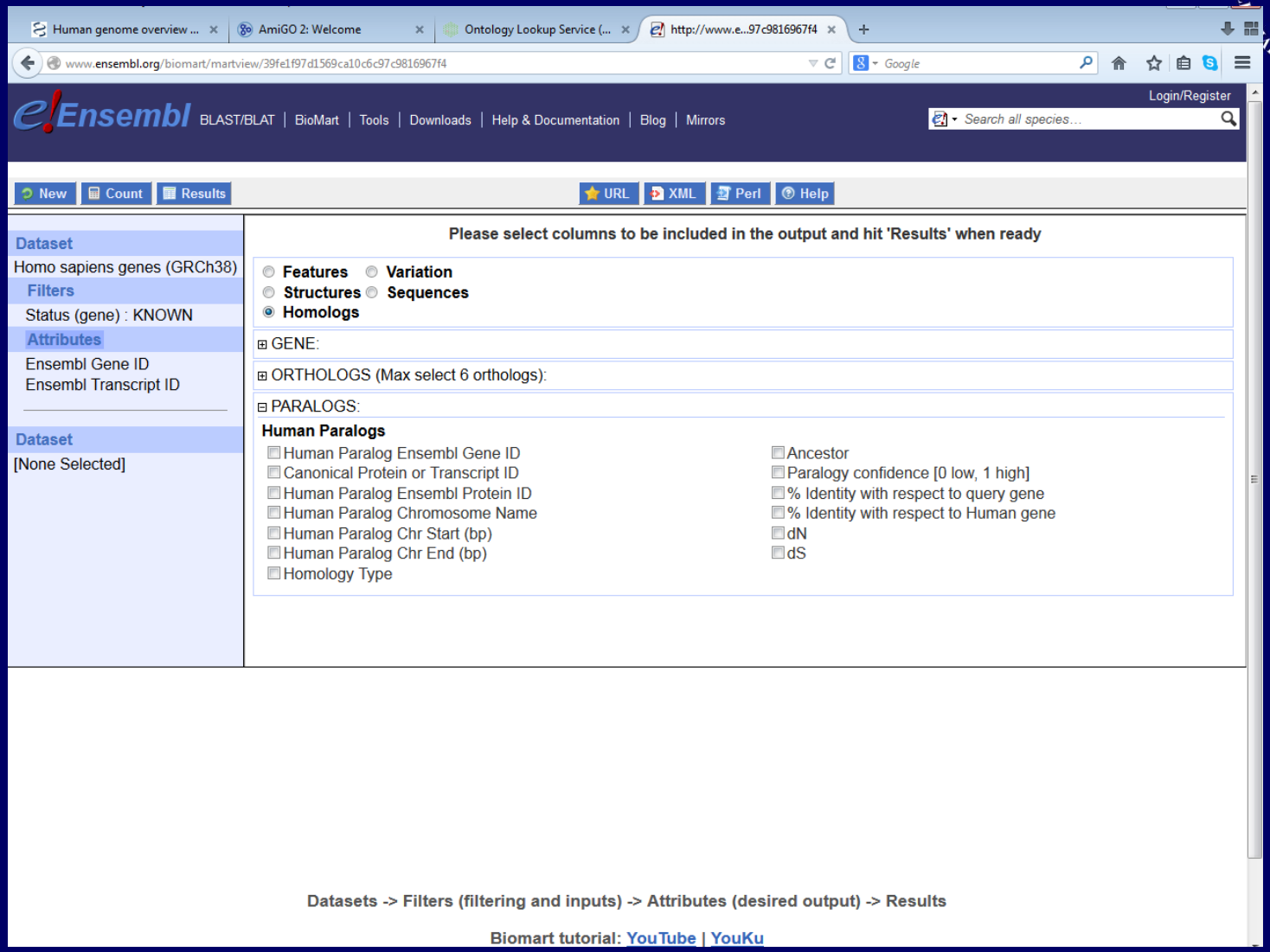

Human genome overview ...
AmiGO 2: Welcome
Ontology Lookup Service (...
http://www.e...97c9816967f4
www.ensembl.org/biomart/martview/39fe1f97d1569ca10c6c97c9816967f4
Ensembl
BLAST/BLAT | BioMart | Tools | Downloads | Help & Documentation | Blog | Mirrors
Login/Register
Search all species...
New
Count
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)
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]
Export all results to
File
TSV
HTML
CSV
TSV
XLS
Unique results only
Go
Email notification to
View
10
rows as
HTML
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 ...
AmiGO 2: Welcome
Ontology Lookup Service (...
http://www.e...97c9816967f4
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
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 ...
AmiGO 2: Welcome
Ontology Lookup Service (...
http://www.e...97c9816967f4
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 Name
GO Term Definition
GO Term Evidence Code
GO domain
GOSlim GOA
GOSlim GOA Accession(s)
GOSlim GOA Description
External References (max 3)
ArrayExpress
ChEMBL ID(s)
Clone based Ensembl gene name
Clone based Ensembl transcript name
Clone based VEGA gene name
Clone based VEGA transcript name
CCDS ID
Database of Aberrant 3' Splice Sites (DBASS3) IDs
DBASS3 Gene Name
Database of Aberrant 5' Splice Sites (DBASS5) IDs
DBASS5 Gene Name
EMBL (Genbank) ID
Ensembl Human Gene IDs
Ensembl Human Transcript IDs
Ensembl Human Translation IDs
LRG to Ensembl link gene
LRG to Ensembl link transcript
EntrezGene ID
EntrezGene transcript name ID
MIM Morbid Accession
MIM Morbid Description
MIM Gene Accession
MIM Gene Description
miRBase Accession(s)
miRBase ID(s)
miRBase transcript name
Protein (Genbank) ID
RefSeq mRNA [e.g. NM_001195597]
RefSeq mRNA predicted [e.g. XM_001125684]
RefSeq ncRNA [e.g. NR_002834]
RefSeq ncRNA predicted [e.g. XR_108264]
RefSeq Protein ID [e.g. NP_001005353]
RefSeq Predicted Protein ID [e.g. XP_001720922]
Rfam ID
Rfam transcript name
UCSC ID
Unigene ID
UniProt/TrEMBL Accession


Human genome overview ...
AmiGO 2: Welcome
Ontology Lookup Service (...
http://www.e...97c9816967f4
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
Variation Source
Variation source description
Variant Alleles
Evidence status
Mapweight
Minor allele
Minor allele frequency
Minor allele count
Clinical significance
Variation Location
Transcript location (bp)
Variation Chromosome Strand
Protein location (aa)
Chromosome Location (bp)
Gene Location and Effect
PolyPhen prediction
PolyPhen score
SIFT prediction
SIFT score
Distance to transcript
CDS Start
CDS End
Protein Allele
Consequence Type (Transcript Variation)
Consequence specific allele
SOMATIC VARIATION INFORMATION:
Datasets -> Filters (filtering and inputs) -> Attributes (desired output) -> Results
Biomart tutorial: YouTube | YouKu


Human genome overview ...
AmiGO 2: Welcome
Ontology Lookup Service (...
http://www.e...97c9816967f4
www.ensembl.org/biomart/martview/39fe1f97d1569ca10c6c97c9816967f4
Google
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
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
Canonical Protein or Transcript ID
Alpaca Ensembl Protein ID
Alpaca Chromosome Name
Alpaca Chromosome Start (bp)
Alpaca Chromosome End (bp)
Homology Type
Ancestor
Orthology confidence [0 low, 1 high]
% Identity with respect to query gene
% Identity with respect to Alpaca gene
dN
dS
Amazon molly Orthologs
Amazon molly Ensembl Gene ID
Canonical Protein or Transcript ID
Amazon molly Ensembl Protein ID
Amazon molly Chromosome Name
Amazon molly Chromosome Start (bp)
Amazon molly Chromosome End (bp)
Homology Type
Ancestor
Orthology confidence [0 low, 1 high]
% Identity with respect to query gene
% Identity with respect to Amazon molly gene
Anole Lizard Orthologs
Anole Lizard Ensembl Gene ID
Canonical Protein or Transcript ID
Anole Lizard Ensembl Protein ID
Anole Lizard Chromosome Name
Anole Lizard Chromosome Start (bp)
Anole Lizard Chromosome End (bp)
Homology Type
Ancestor
Orthology confidence [0 low, 1 high]
% Identity with respect to query gene
% Identity with respect to Anole Lizard gene
dN
dS
Armadillo Orthologs
Armadillo Ensembl Gene ID
Canonical Protein or Transcript ID
Ancestor
Orthology confidence [0 low, 1 high]


Human genome overview ...
AmiGO 2: Welcome
Ontology Lookup Service (...
http://www.e...97c9816967f4
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):
PARALOGS:
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
Datasets -> Filters (filtering and inputs) -> Attributes (desired output) -> Results
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 ... | AmiGO 2: Welcome | Ontology Lookup Service (... | Ensembl: Downloads

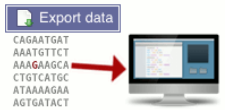
www.ensembl.org/downloads.html

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

Login/Register

Search all species...

Download a sequence or region



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

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

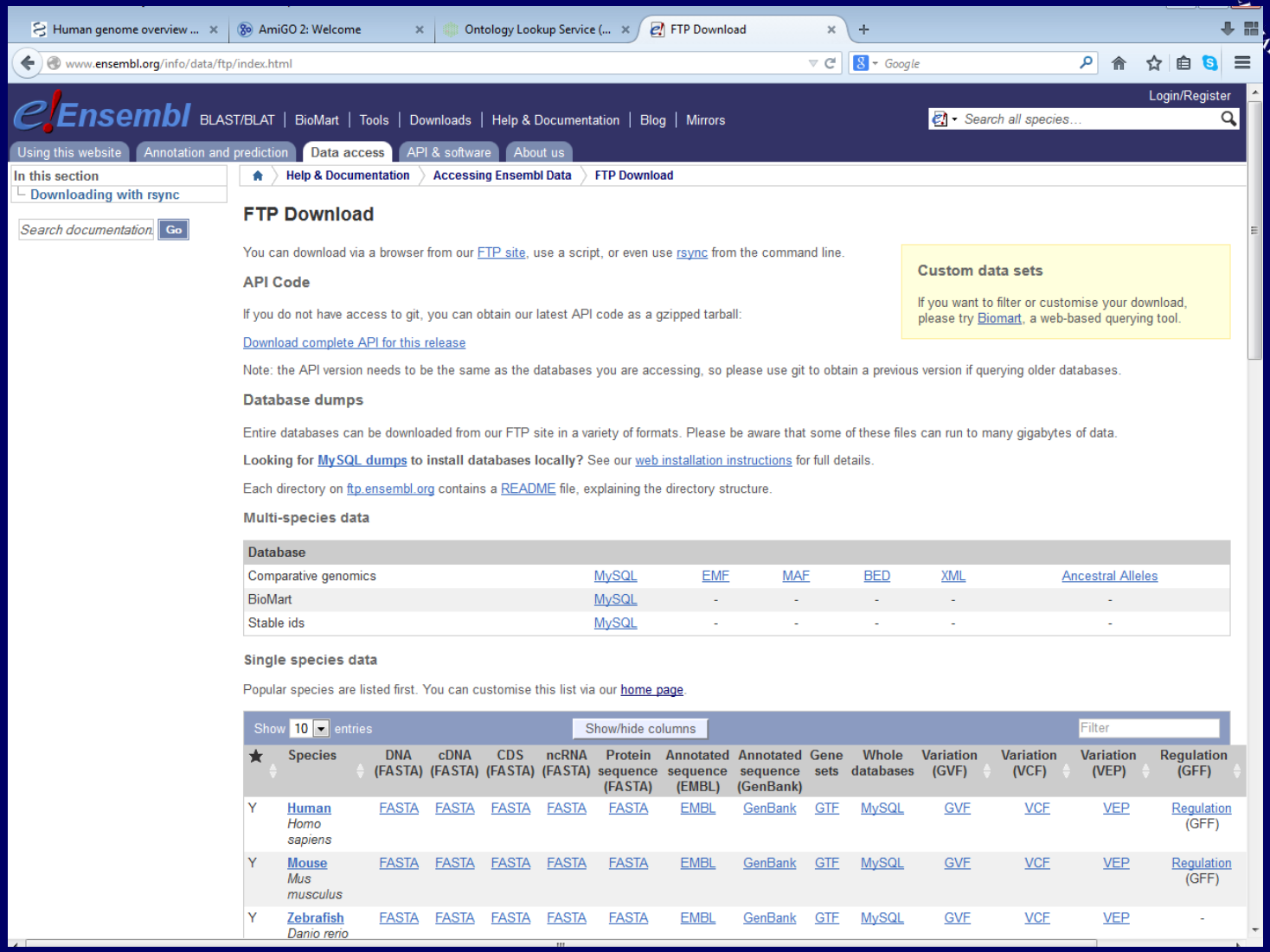
All of our data and software, including pipelines and web code, is available free.

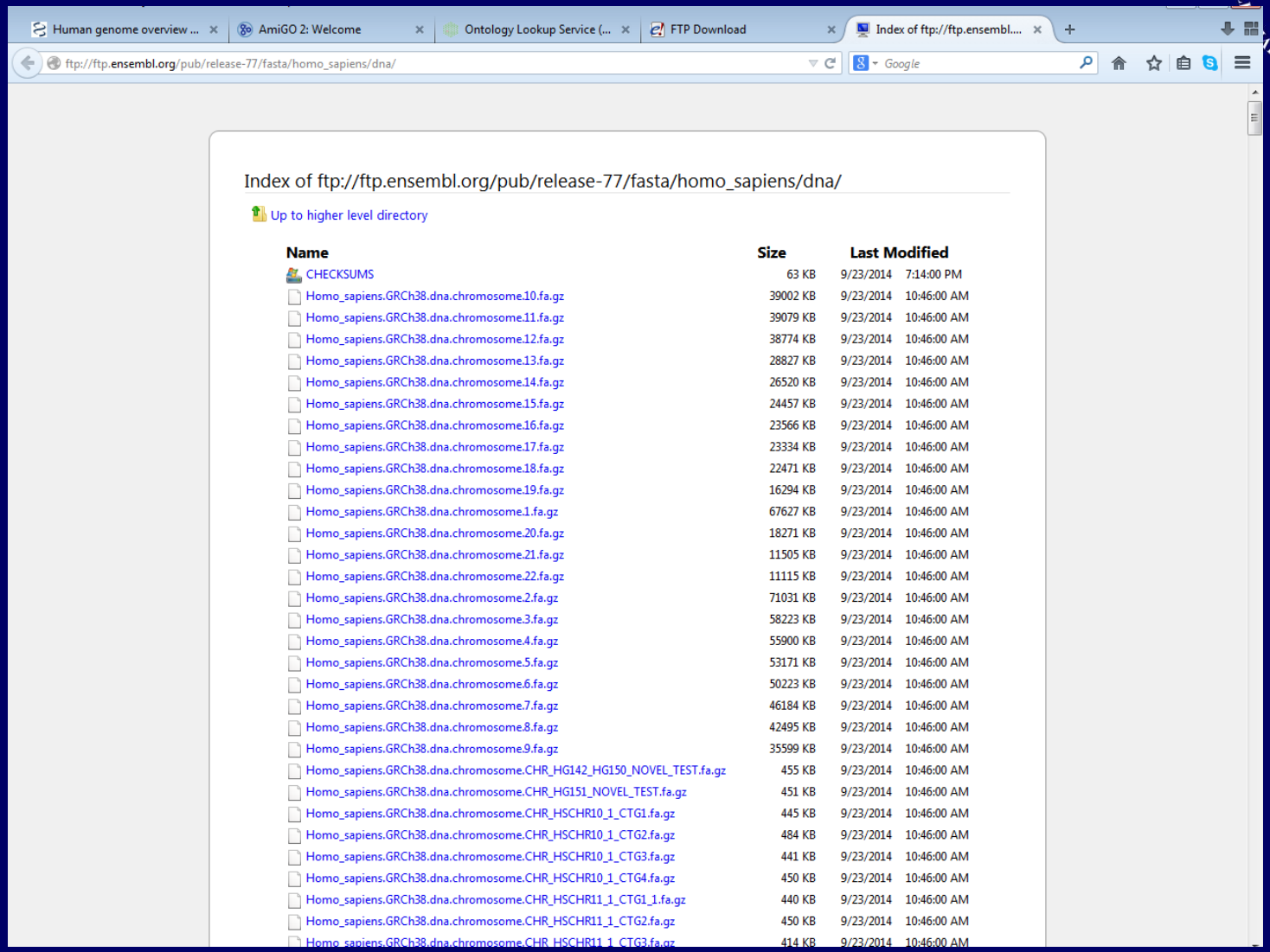
- Download data via FTP
- Ensembl pipeline on GitHub
- Set up your own Ensembl website

Ensembl release 77 - October 2014 © WTSI / EBI

Permanent link

About Ensembl | Privacy Policy | Contact Us

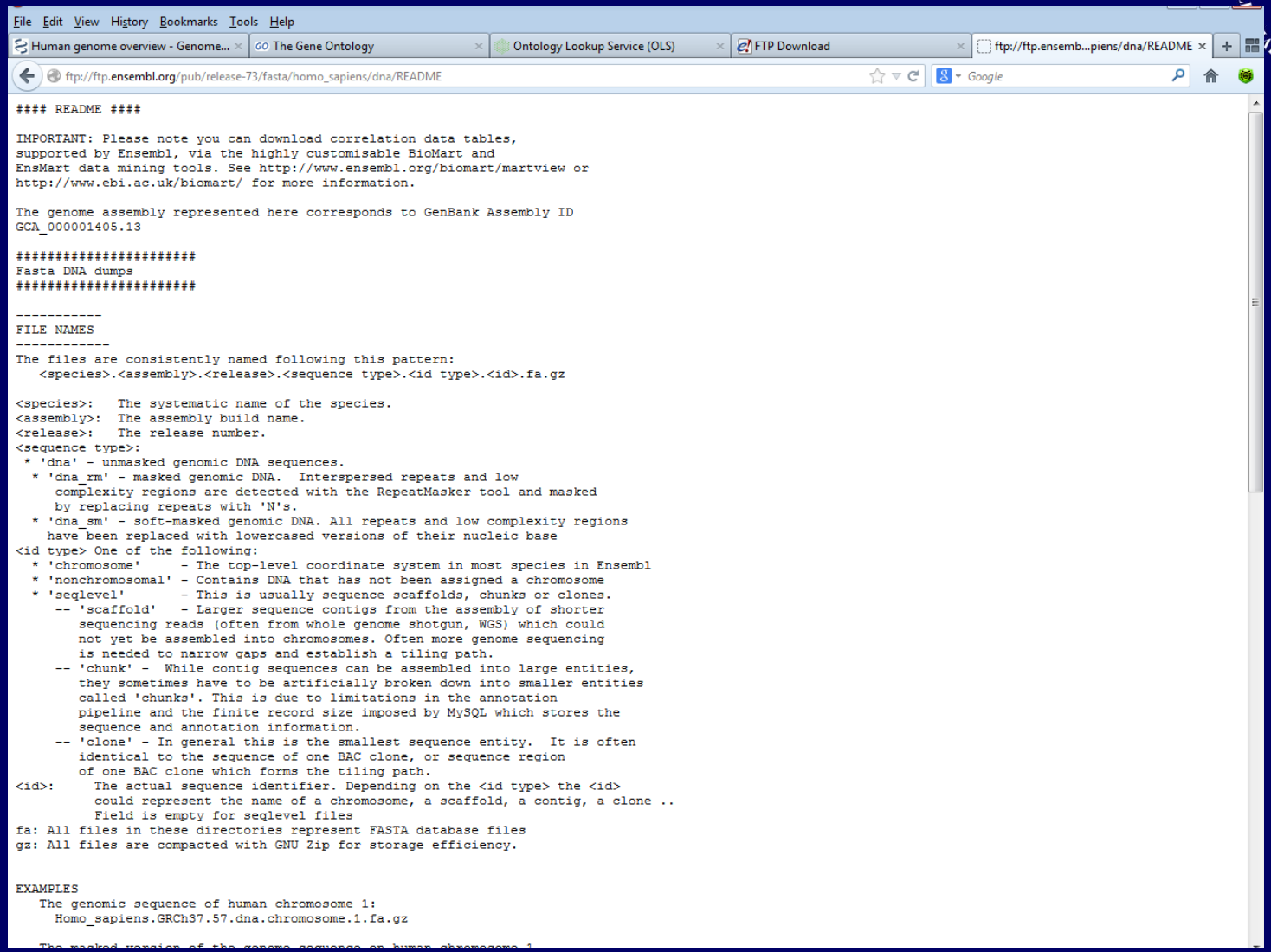

Human genome overview ...
AmiGO 2: Welcome
Ontology Lookup Service (...
FTP Download
www.ensembl.org/info/data/ftp/index.html
Ensembl
BLAST/BLAT | BioMart | Tools | Downloads | Help & Documentation | Blog | Mirrors
Login/Register
Search all species...
Using this website | Annotation and prediction | Data access | API & software | About us
In this section
Downloading with rsync
Search documentation
Go
Help & Documentation › Accessing Ensembl Data › FTP Download
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.
Custom data sets
If you want to filter or customise your download, please try Biomart, a web-based querying tool.
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
Comparative genomics MySQL EMF MAF BED XML Ancestral Alleles
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 Homo sapiens FASTA FASTA FASTA FASTA FASTA EMBL GenBank GTF MySQL GVF VCF VEP Regulation (GFF)
Y Mouse Mus musculus FASTA FASTA FASTA FASTA FASTA EMBL GenBank GTF MySQL GVF VCF VEP Regulation (GFF)
Y Zebrafish Danio rerio FASTA FASTA FASTA FASTA FASTA EMBL GenBank GTF MySQL GVF VCF VEP -



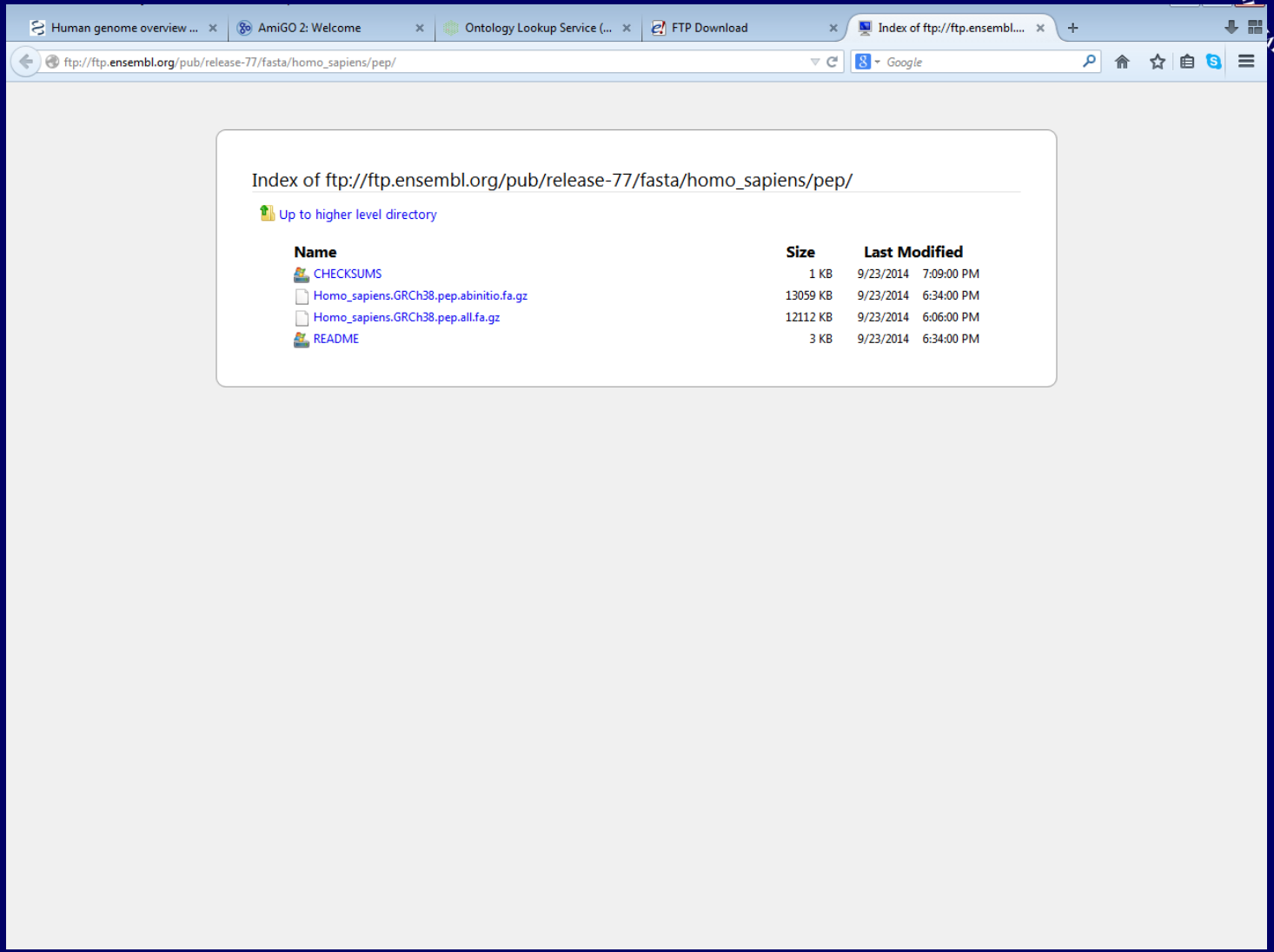
Index of ftp://ftp.ensembl.org/pub/release-77/fasta/homo_sapiens/dna/

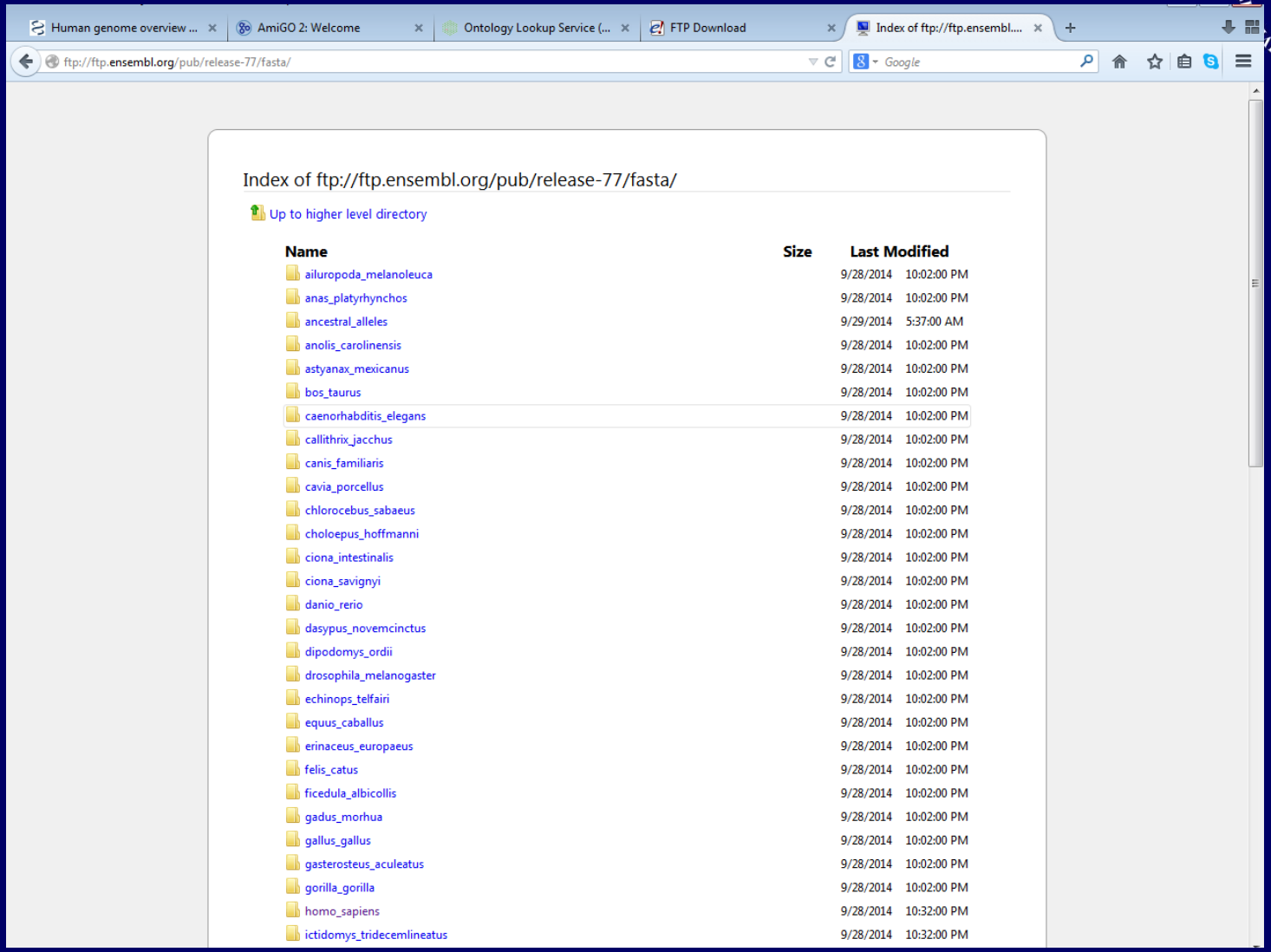
Up to higher level directory

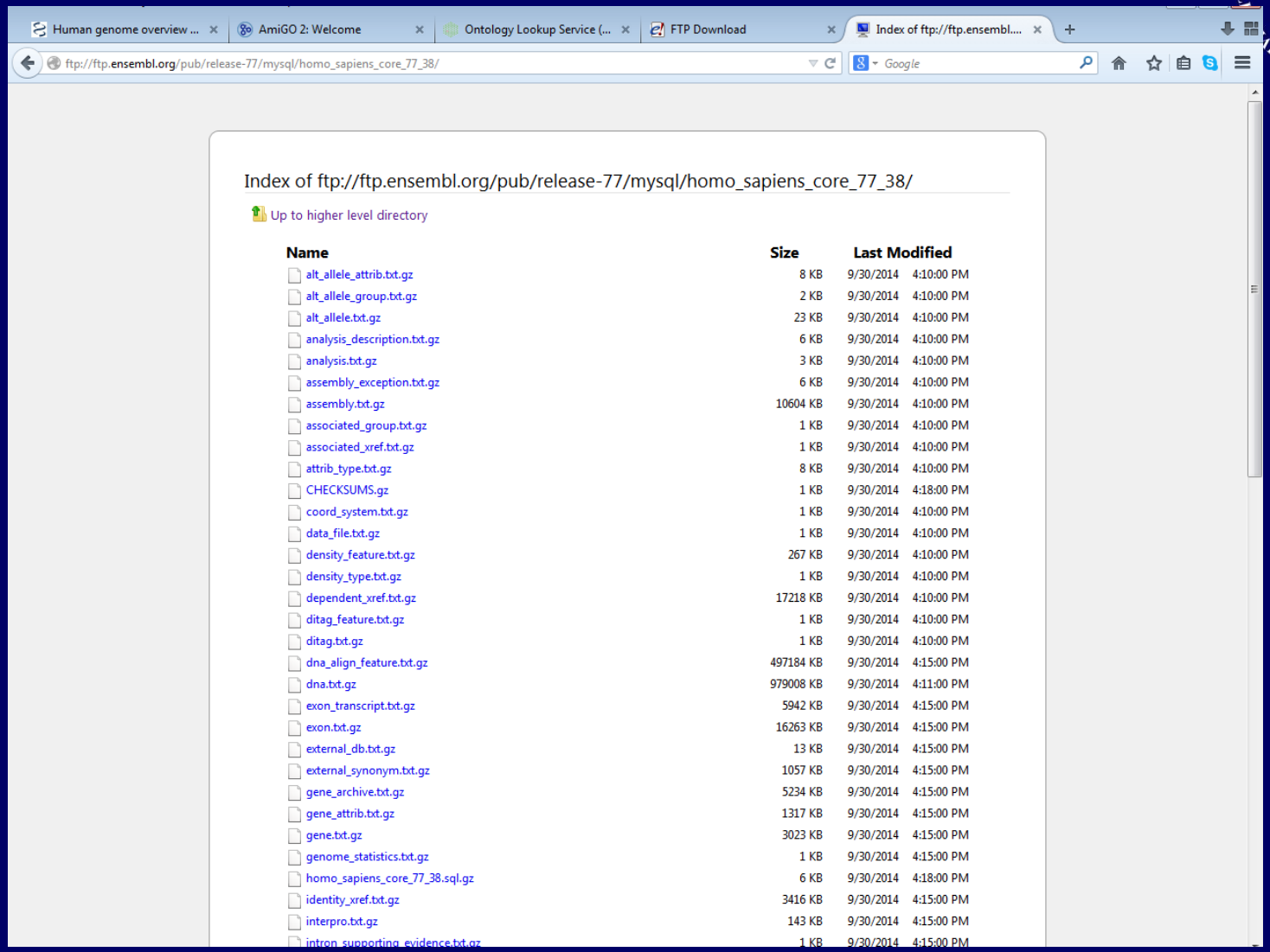
Name	Size	Last Modified
CHECKSUMS	63 KB	9/23/2014 7:14:00 PM
Homo_sapiens.GRCh38.dna.chromosome.10.fa.gz	39002 KB	9/23/2014 10:46:00 AM
Homo_sapiens.GRCh38.dna.chromosome.11.fa.gz	39079 KB	9/23/2014 10:46:00 AM
Homo_sapiens.GRCh38.dna.chromosome.12.fa.gz	38774 KB	9/23/2014 10:46:00 AM
Homo_sapiens.GRCh38.dna.chromosome.13.fa.gz	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_HSCHR10_1_CTG1.fa.gz	445 KB	9/23/2014 10:46:00 AM
Homo_sapiens.GRCh38.dna.chromosome.CHR_HSCHR10_1_CTG2.fa.gz	484 KB	9/23/2014 10:46:00 AM
Homo_sapiens.GRCh38.dna.chromosome.CHR_HSCHR10_1_CTG3.fa.gz	441 KB	9/23/2014 10:46:00 AM
Homo_sapiens.GRCh38.dna.chromosome.CHR_HSCHR10_1_CTG4.fa.gz	450 KB	9/23/2014 10:46:00 AM
Homo_sapiens.GRCh38.dna.chromosome.CHR_HSCHR11_1_CTG1_1.fa.gz	440 KB	9/23/2014 10:46:00 AM
Homo_sapiens.GRCh38.dna.chromosome.CHR_HSCHR11_1_CTG2.fa.gz	450 KB	9/23/2014 10:46:00 AM
Homo_sapiens.GRCh38.dna.chromosome.CHR_HSCHR11_1_CTG3.fa.gz	414 KB	9/23/2014 10:46:00 AM




```
#### README ####

IMPORTANT: Please note you can download correlation data tables,
supported by Ensembl, via the highly customisable BioMart and
EnsMart 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
```


Human genome overview ...
AmiGO 2: Welcome
Ontology Lookup Service (...
FTP Download
Index of ftp://ftp.ensembl....
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 ...
AmiGO 2: Welcome
Ontology Lookup Service (...
FTP Download
Index of ftp://ftp.ensembl....
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
choloepus_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
dasypus_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



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_attrib.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_attrib.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.