Browsing Genomes with Ensembl



www.ensembl.org

www.ensemblgenomes.org

 \odot \odot

Contents

- Finding information about species and genomes in Ensembl, Demo
- Exploring genomic regions in Ensembl, Demo
- Human genes and transcripts in Ensembl, Demo
- Exploring variants in Ensembl, Demo
- <u>Annotating genetic variants with the VEP, Demo</u>
- <u>Gene trees and homologues, Demo</u>
- <u>Whole genome alignments, Demo</u>
- Features that regulate gene expression, Demo
- Bulk export of data with BioMart, Demo
- <u>Custom data, Demo</u>

Finding information about species and genomes in Ensembl, Demo

Demo: Introduction to Ensembl

Ensembl

Homepage

The front page of Ensembl is found at <u>ensembl.org</u>. It contains lots of information and links to help you navigate Ensembl:

	=P Tools BioMart Downloads Help & Docs Blog	Login/Registe
Link back to BioMart > All tools Export purpor datasets f Ensembly this data-m tool Ensem	blastication blastication Blastication Blastication Blastication Blastication Blastication Blastication Blastication Blastication Blastication Blastication Blastication Blastication Blastication Blastication Blastication Blastication Blastication Analyse your own varia Blastication Analyse your own varia Blastication Analyse your own varia Blastication Analyse your own varia Blastication Blas	Ensembl is a genome creating of the period sector (VEP) for all supported species. Ensembl Bleeses 104 (May 2021)
Search All species	▼] for	Optime to a present carbona water of the set of th
e.g. BRCA2 or ra	Go 5:62797383-6362766 or rs699 or coronary heart disease	and what is a new
All genomes Select a species	Search Favourite genomes	New assemblies with gene and protein annotation every two weeks. Note: species that already exist on this site will continue to be updated with the full range of annotations.
Pig breeds Pig reference genome and 12 aptiliti Prop-down lis	and breeds SIII using OBCh322 t of of species	The Ensembl Rapid Release website rovides anotation for recently produced, publicly available verobrate and non-vertexina genomes from biodiversity initializes such as Raviw Tras of Life the Durdent Revenues for the the theory of the second second Durdent Revenues for the theory of the second se
View full list of all species	Zebrafish GRCz11	on Rapid Release news or our blog en Rapid Release news or our blog • 08 Sep 2021: Job: Software Developer of
		■ 23 Aug 2021: Job: Outreach Officer r

On the right-hand panel you can see the current release number and what has come out in this release. To access old releases, scroll to the bottom of the page and click on **View in archive site** in the right-hand corner.

View in archive site	
Search	The following archives are available for this page:
View in archive site Search ⊟ Help topics ⊢ Frequently Asked Questions Video Tutorials Glossary Contact HelpDesk	The following archives are available for this page: • Ensembl GRCh37: Full Feb 2014 archive with BLAST, VEP and BioMart • Ensembl 110: Jul 2023 (GRCh38.p14) - patched/updated gene set Mar 2023 • Ensembl 109: Feb 2023 (GRCh38.p13) - patched/updated gene set Nov 2022 • Ensembl 108: Oct 2022 (GRCh38.p13) - patched/updated gene set Jul 2022 • Ensembl 107: Jul 2022 (GRCh38.p13) - patched/updated gene set Apr 2022 • Ensembl 106: Apr 2022 (GRCh38.p13) - patched/updated gene set Nov 2021 • Ensembl 105: Dec 2021 (GRCh38.p13) - patched/updated gene set Aug 2021 • Ensembl 105: Dec 2021 (GRCh38.p13) - patched/updated gene set Mar 2021 • Ensembl 104: May 2021 (GRCh38.p13) - patched/updated gene set Mar 2021 • Ensembl 103: Feb 2021 (GRCh38.p13) - patched/updated gene set Aug 2020
	 Ensembl 102: Nov 2020 (GRCh38.p13) - patched/updated gene set Sep 2020 Ensembl 101: Aug 2020 (GRCh38.p13) - patched/updated gene set Mar 2020 Ensembl 100: Apr 2020 (GRCh38.p13) - patched/updated gene set Jun 2019 Ensembl 99: Jan 2020 (GRCh38.p13) - patched/updated gene set Aug 2019 Ensembl 98: Sep 2019 (GRCh38.p13) - patched/updated gene set Jun 2019 Ensembl 98: Sep 2019 (GRCh38.p12) - patched/updated gene set Mar 2019 Ensembl 97: Jul 2019 (GRCh38.p12) - patched/updated gene set Mar 2019 Ensembl 96: Apr 2019 (GRCh38.p12) - patched/updated gene set Nov 2018 Ensembl 95: Jan 2019 (GRCh38.p12) - patched/updated gene set Jul 2018 Ensembl 95: Jan 2019 (GRCh38.p12) - patched/updated gene set Jul 2018 Ensembl 97: Oct 2014 (GRCh38.p2) - patched/updated gene set Aug 2014 Ensembl 77: Oct 2014 (GRCh38) - patched/updated gene set Sep 2013 Ensembl 54: May 2009 (NCBI 36) - patched/updated gene set Oct 2008

Click on the links to go to the archives. Alternatively, you can jump quickly to the correct release by adding **e** plus the **release number** in the URL. For example <u>e98.ensembl.org</u> jumps to Ensembl release 98.

Available species

Scroll back up to the top of the homepage. You can view all available species by clicking the **View full list of all species** link underneath the coloured search block.

You can search for your species of interest (either the common or scientific name) using the search bar at the top right-hand corner of the table. Click on the common name of your species of interest to go to the species information page. We'll click on **Human**.

👔 Human (GRCh38.p13) 🔻			
Search Human (Homo sapiens) Search all catogories • Search Go e.g. BRCA2 or 17:65962802-64038237 or re509 or osteoarthritis			
Genome assembly: GRCh38.p13 (GCA_000001405.28) Search More information and statistics Download DNA sequence (FASTA) Convert your data to GRCh3028 octimates Differ assemblies GRCh37 Full Feb 2014 archives with BLAST, VEP and Booklet 2015	View karyotype	Gene annotation What can I find? Protein-coding and non-coding genes, splice variants, cDNA and protein sequences, non-coding RNAs. Image: A sequence of the se	Example transcript
Comparative genomics What can I find? Homologues, gene trees, and whole genome alignments across multiple genome. 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 observations of the second variation in Ensempt More about variation in Ensempt Cogenities all variants (GVP)	ATCGAGCT ATCCAGCT ATCGAGAT Example variant
Regulation What can I find? DNA methylation, transcription factor binding sites, histone modifications, and regulatory features such as orhancers and repressors, and microarray annotation If More about the Ensembli regulatory build and microarray annotation Ensemblinemental data sources Diameted all regulatory (Setting 100FF)	Example regulatory	eatures in Ensembl	Example phenotype

Species information

Here you can see links to example features and to download flatfiles. To find out more about the genome assembly and genebuild, click on **More information and statistics** under the **Genome assembly** section.

🙀 Human (GRCh38.p13) 🕶		
Real Human assembly and gene annotation		
Assembly	Statistics	
This site provides a data set based on the December 2013 Homo sapiens high coverage assembly GRCh38 from the Gamma Reference Consulting 6. This assembly and by UCSC to create their ho38 distabase. The data set consists	Summary	
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.	Assembly	GRCh38.p13 (Genome Sence Consortium Human Build 38), INSDC Assembly GCA_000001405.28.67, Deb. 13
This release of the assembly has no ploying properties.	Base Pairs	3,096,649,726
· contig length total 3.4 Gb. Description of the geno	Path Length	"Ctaticting about the
 chromosome length total 3.1 Gb (excluding hapiotypes). 	Assembly provider	Entre Contraction and the
It also includes 281 at ioci aceffolds, mainly in the LRC/KIR complex on compositions) and the MHC are of the provided of the complex of the provided of the complex of the	And othing method	Full genebuild
Watch a video on YouTuber@ about patches and haplotypes in the Human genome.	Genebuild released	genome and gene
	Genebuild last updated/patched	Mar 2021
Patches brocecc	Database version	104.38
As the GRC maintains and improves the assembly, patches are being introduced. Currently, assembly patches are of two	Gencode version	• annotation
Abea:		

Here you'll find a detailed description of how to the genome was produced and links to the original source. You will also see details of how the genes were annotated.

The current genome assembly for human is GRCh38. If you want to see the previous assembly, GRCh37, visit our dedicated site <u>grch37.ensembl.org</u>.

GRCh37 BLAST/BLAT VEP Tools BioMart Downloads H	letp & Docs	arch all species Q
Search All species v for e.g. BRCA2 or human 5:62797365-63627669 o	(Co) re009 or coronary heart disease	About this archive This archive is based on Ensembl Release 75 data, and gives continuing costs to human ancembly GRCh37. Human variation and regulation data as since bern updated in April 2021. AyGCL dumps of human distabases on the most recent schema version re available on our ETP ass.
Browse a Genome The Ensembl project produces genome databases for vertebrates and other eukaryotic spo Available genomes Human GRCN37 p13 Levra m	I Sies, and makes this information freely available online. I Use GRCh387 User	Ensembl GRCh37 Release 104 (May 2021) Updated regulatory build. Updated variation data including the latest data from dbSNP build 154, Clinthar and COSMIC. More releases news of on our blog

Ensembl Genomes

Homepage

Let's take a look at the Ensembl Genomes homepage at ensemblgenomes.org.

<i>C</i> EnsemblGenomes Providing genome data for no	in-vertebrate species, with tools for th	ne manipulation, analysis and visualisation of that d	lata	Contact us
Latest release notes, updates & news from our blog		Search all genomes		60
Clensembl COVD-19			urch for	a genome
SARS-CoV-2 Genome sequen	ce & annota	Triticum aestivum IWGSC		
C. Ensembl Rupid Rolesse	COVTD	Arabidoosis thatiana		
2-weekly releases of new assemblies with gene & protein feature annotation	S	Go to Ensembl Plant		
C. EnsemblM	letazoparid D	e!EnsemblProt	tipts , ka ta	toxo
Caenorhabditis elegans Webce/235	Каріа К	Plasmodium fakiparum 3D7 ASM276v2	LIMKS CO	
Drosophia melanogaster BDGP6.28 Bombyx mori	browser	Dickyastelium discoldeum dicty_2.7	specific .	sites
ASM15162v1		ASM14294v1		
Go to Ensembl Me	Mazoa	Go to Ensembl Protist	ts	

Click on the different taxa to see their homepages. Each one has a different colour-coding, but they are all structured in a similar format to the Ensembl main site.

CEsseenbilWants- march	8.457 Boller Tools Granisans Parpil	Laganhyan Ina I Bay	C ExceeduPycelate : mage surr build tue builds te	Lago Papato No 1 Pag (En Santa Factor - Kator	C Ensemblikingi - math buti bass bas bassa bas	ion i ing Contraction (
teater of games and formage consult	•)v	When a constraints transition has not a particular anomaly non-re-bible deckspin p_1 , including a the bible deckspin p_2 are subset to deck and a constraint of a of bibles to a subset of the deckspin p_2 and the deck constraints of the deck to deck the deck of the deck to deck the deck of the de	Smith of partial - Jor - Jor - Jor	Pharts New in Nations (1) * Distance (2)	National Conference of Confere	Venice New In Address 12 + Similar - Analysis of the Similar Similar Similar + No New or operating periods 11% bet Similar
Nigerana - Anne a sanne	Analysis of the second se	 A standard barr dark for standard s	With the second seco	extension of a price service and a constraint of a constr	V granden Mark Andream Construction Mark Andre	E E
	Marked Algorithm	Active the formation of the contract of the termination of the two of the formation of the two of the two of the termination of the two of two of the two of tw	And the second s	Part Management P or day	Australia generative Alternative	and the second s
		Anna Casard Charlen an anna Anna Charl C Ensemblikelezze - mach (batt (bane (bas) borow) m	tapahayan périma jing 👔 Jacor Parch Haras	• New (, Series V) address and and a series of • New (, Series V) address and and a series (series (series) series) series (series	tegelingete tege	¹⁴ Sector Sector P. C. Statistics and Rev. Phys. Rev. Biol. 2014; Sector P. S. Statistics and S. S. S. Sector P. Sec
		National and a second and a sec	Phatis Mars in Platases 11 4: Spanis Mar 4: Spanis Mars 4: Spanis 4: Spanis Mars 4: Spanis Mars 4: Spanis 4: Span	Baset for a general second access. Baset for a general for gene for second access. Applied access Baset gene for second access Applied access Baset for a general for second access	And the state Transmission of the second second second second at the interaction of the second second second second second is the second second second second second second second is the second second second second second second second second second is the second seco	
		Land Landa - V Constants angus Sub-Constants angus Sub-Constants Constants angus Sub-Constants angus Sub-C	Angeles delesas (2012), 4 Angelesas (2012), 4 Angeles		Council Backets Unset Backets I have a solutione to hadren and antique particle. You an inset further attracts to all content of the URL anticipation of approx. The all content of the URL anticipation of approx.	
		economic antinopolitic technology (Conomic antinopolitic technology (Conomic antinopolitic) technology (Conomic antinopolitic) technol	 Kapiten antip Jani; Ropite celastro de ANIII Ropite celastro de ANIIII Ropite celastro de ANIIIII Ropite celastro de ANIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII	There is a descent set on the original set of the	Encode consideration and the filter of provide provide constraints of provide constraints o	
		Review of the second se	A regimes: related a constrainty A regimes: related a constrainty A regimes: related a relation of the relation o	EXE Exect of an encoder and a financial rate of the second seco	1000 COL (1000)	

You can navigate most of the taxa in the same way as you would with Ensembl.

Ensembl Bacteria

Ensembl Bacteria has a large number of genomes and has a slightly different method to the other Ensembl sites. Let's look at it in more detail.

<image/>					Login/Register
<section-header> Search for a gene • or nis2 or undere • or nis2 or undere</section-header>	🛃 EnsemblBacteria 🖌 hmmer blas	F Tools Downloads Help & Docs	Blog	Tearch Ensembl Bacteria	Q,
<section-header> Search for a gene Search for a gene Search and agencies. I g its or undime Search for a geneme for active a second for active</section-header>					
 A total of 31,332 bacterial and archaeal genomes A total of 31,332 bacterial and archaeal genomes A total of 31,332 bacterial and archaeal genomes 	Search for a gene Search for a gene Search for a gene Search for a gene of the start type of the search of the sea	pr a genome ing the name of a generation is to find Escherichia affer into the search Search	Archive sites The following archive sites are available to access a Release 49, December 2020 <u>eg49-bacteria er</u> Release 45, September 2019 <u>eg45-bacteria er</u> Felesase 40, July 2018 eg40 <u>bacteria ers</u> FOR Release 47, General Control Control era ense	previous versions of data: teambillarg & teambillarg & argiff mbillarg &	
What's New in Release 51 Release 51 of Ensembl Bacteria has no major updates since the previous release. As for release in our blog St. • Genomes • A total of 31,332 bacterial and archaeai genomes To genomes data from any	Find a genome - click in the 'browse a genome' box abov name to find matching genomes Yeev full ist of all Ensembl Bacteria secsies Access Ensembl Bacteria strogrammatically	and start typing your genome	Ensembl Bacteria Ensembl Bacteria is a browser for bacterial and arc of the International Nucleotida Sequence Database the EBI, GenBark at the NCP and the DNA Datab Data access	haeal genomes. These are taken from the <u>Colaboration</u> #, the European Nucleotid ase of Japan).	e databases e Archive at
	What's New in Release 51 Release 51 of Ensembl Bacteria has no major updates since previous release. As for release 49, we only represent non- rodundant bacterial genomes as defined by criteria set out by UniPort. See mere details about this update in our <u>blog</u> s. • Genomes • A total of 31,332 bacterial and archaeal genomes	e Did you know? eREST To access Ensembl Genomes data from any	Data can be visualised through the Exembl genom Peri and RESTMA APProtate spice accessible thro containing tild ata advances to the spice of the 100 key bacterial genomes have been classified into families using H genomes have been classified into families using H Ensembl Genomes Eusepeuble	e browser and accessed programmatical gen public MySGL databases and our FT ELOIDE and Company and genes for AMAP and PANTHER more details.	lly via our 'P site n of over om all are system
Data Programming language, try our REST ansolation of pathogen-host interaction data (PH/base % version 2019-09-16 Alignments to Riam & Covariance models (Riam 12.2) visible in separate track (Pfan models) Updated protein features of all species using InterProtSean & version 81	Data Annotation of pathcgen-host interaction data (PHI-ba version 2019-09-16 Alignments to <u>Rfams</u> covariance models (Rfam 12.2 visible in separate track ("Rfam models") Updated protein features for all species using <u>InterProScan</u> & version 81	Programming language, try our REST services ¹⁷ . For full documentation, including examples from a wide range of language, visit http://rest.ensembl.org.i5	for the analysis and visualisation of genomic data. I	ior details of our funding please click here	b The second sec

There's no drop-down species list for bacteria as it would be hard to navigate with the number of species. You can click the **View full list of all Ensembl Bacteria species** link underneath the coloured search block. Search for your species of interest using the filter in the top right-hand corner of the table.

-1					Login/Register
🥑 EnsemblBacteria 🗸 🛛 🕺 HMMER BLAST Tool	s Downloads Help 8	Docs Blog		很 • Search Ensei	mbl Bacteria Q
Find a Species					
Show 25 🗸 entries				Filter Clostridioide	s difficile
Species	Assembly	Taxonomy ID	Serotype	Publications	Present in
	*				pan-taxonomic compara
Clostridioides difficile 630 (GCA_000009205)	ASM920v2	272563			Y
Clostridioides difficile ATCC 9689 = DSM 1296 (GCA_001077535)	ASM107753v2	1121308			N
Clostridioides difficile CD160 (GCA_000449425)	ASM44942v2	1151292			N
Clostridioides difficile CD196 (GCA_000085225)	ASM8522v1	645462			N
Clostridioides difficile F501 (GCA 000450805)	ASM45080v2	1151372			N
Clostridioides difficile str. RA09 70 (GCA 001299495)	ASM129949v1	1496			N
Clostridioides difficile (GCA_002301955)	ASM230195v1	1496			N
Clostridioides difficile str. CD-B18-123 (GCA_005502205)	ASM550220v1	1496			N
Clostridicides difficile str. WCH065050 (GCA_005786645)	ASM578664v1	1496			N

Alternatively, you can find a species by typing the species name into the **Search for a genome** search box at the top of the page. A drop-down list will appear with any species matching the name you entered.

For example, to find a sub-strain of *Clostridioides difficile* start typing in the species name. Due to the autocomplete, you'll see useful results as soon as you get to *Clostridio*.

Search for a gene		Search for a genome	Archive sites
Search all species	Go	clostridio	The following archive sites are a
e.g. ftsZ or uridine*		Clostridioides difficile (GCA_002301955), (TaxID 1496)	Π.
-		Clostridioides difficile (GCA_900074795), (TaxID 1496)	
		Clostridioides difficile (GCA_900164675), (TaxID 1496)	
		Clostridioides difficile (GCA_900684045), (TaxID 1496)	
Search for a gene - type th	e name of a gene or	Clostridioides difficile (GCA_900686955), (TaxID 1496)	
above		Clostridioides difficile (GCA_901002355), (TaxID 1496)	
Find a genome - click in the	e 'browse a genome'	Clostridioides difficile (GCA_901005725), (TaxID 1496)	
name to find matching gen	omes	Clostridioides difficile (GCA_901006595), (TaxID 1496)	
View full list of all Ensemble	Bacteria species	Clostridioides difficile (GCA_901006875), (TaxID 1496)	
 Access Ensembl Bacteria r 	programmatically	Clostridioides difficile 630 (GCA_000009205), (TaxID 2	72563)
	<u>grannanouny</u>		of the International Nucleotide Se

The drop down contains various strains of *C. difficile*. Let's choose *C. difficile* 630. This will take us to another species information page, where we can explore various features.



Unlike the *Homo sapiens* species information page, there is no prose description of the genome or gene annotation, as these pages were generated automatically.

Ensembl Rapid Release

Our newest genomes, such as those coming from the <u>Darwin Tree of Life</u>, are available <u>rapid.ensembl.org</u> with limited annotation.

ools	BLAST >		Latest Genomes
ll toole	Search our genomes for your DNA		We have 4 new genomes this release:
10010	or protein sequence		Cimex lectularius (Bed bug) - GCA_000648675.3 - Harlan Olive haboon
			Sumatran orangutan - General/2880775.3
	Search		 Salmo salar (Atlantic salmon) - 0.5A_905237065.2
	Comathunshus populus /	2molt as for	View all species and download data
	Cananynonus parvalus (Go	Alawart agus ag
	e a Camarbunchus naneil	us 2:361680-384534 or Clutia hemisphaerica IRR001650	Ivewes' genomes
	old: committy total barran		—
sembl Rapid Rel kaning we can ma	lease is a new site designed to make our ake our gene sets available with minimal	data available more quickly. Release of data occurs on a two-week cycle, delay once the annotation is complete. For each species we provide a gene se	a
neembl Rapid Rel leaning we can mi long with additiona is important to no urrently we do not <i>le</i> are however wo lore details about	ease is a new site designed to make our ake our gene sets available with minimal al features such as protein feature anear to that Ensemb Rapid Release is by nat provide homology data, gene symbol as ming on adding more functionality over Ensemb! Rapid Release and the const	data available more quickly. Release of data occurs on a two-week cycle, delay once the annotation is complete. For each species we provide a gene se and BLAST functionality. The onlys fully detuned as a typical data release on <u>www.ensembi.org</u> . algom in (which uses homology data), data archiving or programmatic access. the coming months to further impreve usability.	и
nsembl Rapid Rel leaning we can m long with additions is important to no urrently we do not fe are however we lore details about Il genomes	ease is a new site designed to make our ake our gene sets available with minimal al features such as protein feature and provide homology data, gene symbol as forging on adding more functionality over Ensemb! Rapid Release and the o	data available more quickly. Release of data occurs on a two-week cycle, delay oroce the annotation is complete. For each species we provide a gene se in due BLAST functionally. The orbus fully featured as a typical data release on <u>www.ensembi.org</u> . aignm in (which uses homology data), data archiving or programmatic access. become months to further improve usability. Douttent Rapid Release Highlighted genomes	и
nsembl Rapid Rel eaning we can m ong with additiona is important to no urrently we do not e are however wo ore details about Il genomes - Select a specie:	ease is a new sile designed to make our ake our gene sets available with minimal al features such as protein feature and provide homology data, gene symbol as right and and more functionality over Ensembl Rapid Release and the case a -	data available more quickly. Release of data occurs on a two-week cycle, offay once the annotation is complete. For each species we provide a gene se on the BLAST functionality. The only of two sets the optical data release on <u>www.ensembi.org</u> 62. aight in (which uses homology data), data archiving or programmatic access. becoming months to further improve usability. Dout the Rapid Release Highlighted genomes CCA approaches 1	и и
sembl Rapid Rel paning we can m ing with additional s important to no irrently we do not a ren however we ore details about I genomes Select a specie	ease is a new sile designed to make our ake our gene sets available with minimal al features such as protein feature and provide homology data, gene symbol as indig on adding more functionality over Ensemb! Rapid Release and the case as -	data available more quickly. Release of data occurs on a two-week cycle, obiay once the annotation is complete. For each species we provide a gene se mont BLAST functionality. The only of the data release on <u>www.ensembl.org</u> 92 signm in (which uses homology data), data archiving or programmatic access. The contegr months to further improve usability. DOUT on Rapid Release Highlighted genomes Camaritynchus parvulus (Smail tree finch) - <u>Camaritynchus parvulus (Vr.1</u>	R
nsembl Rapid Rel eaning we can m ong with addition: is important to no <i>le</i> are however we tore details about all genomes – Select a specie: iew and download	ease is a new sile designed to make our ake our gene sets available with minimal al features such as protein feature and provide homology data, gene symbol as noking on adding more functionality over Ensembl Rapid Release and the s s	data available more quickly. Release of data occurs on a two-week cycle. delay once the annotation is complete. For each species we provide a gene se mod BLAST functionality. re not not sully delay data release on <u>yew ensembling</u> of signment (which uses brondogy data), data archiving or programmatic access. the complete morths to further improve usability. Source the Republic Release Highlighted genomes Examplify Camarhynchus parvulus (Small tree finch) - CCA_902806525.1 Camarhynchus_parvulus.(V1.1	x

Exploring genomic regions in Ensembl, Demo

Start at the Ensembl front page, <u>ensembl.org</u>. You can search for a region by typing it into a search box, but you have to specify the species.

To bypass the text search, you need to input your region coordinates in the correct format, which is **chromosome, colon, start coordinate, dash, end coordinate**, with no spaces for example: **human 4:122868000-122946000**. Type (or copy and paste) these coordinates into either search box.

Search	
All species v for	
chicken 4:53544500-53598000 Go	
e.g. BRCA2 or rat 5:62797383-63627669 or rs699 or coronary heart disease	or
≦ Chicken (bGalGal1.mat.broiler.GRCg7b) ▼	
Search Chicken (Gallus gallus)	
Search Chicken (Gallus gallus) Search all categories 4:53544500-53598000	Go

Press Enter or click Go to jump directly to the Region in detail Page.



Click on the [®] button to view page-specific help. The help pages provide text, labelled images and, in some cases, help videos to describe what you can see on the page and how to interact with it.

The Region in detail page is made up of three images, let's look at each one in detail.

The first image shows the chromosome:

Chromosome 4: 122,868,000-122,946,000

Assembly exceptions chromosome 4	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1		1011 101 1011		
Partine grander and a second	Cytogenetic bands	Haplotypes	Our region	of inter	est

The region we're looking at is highlighted on the chromosome. You can jump to a different region by dragging out a box in this image. Drag out a box on the chromosome, a pop-up menu will appear.



If you wanted to move to the region, you could click on *Jump to region (### bp)*. If you wanted to highlight it, click on *Mark region (###bp)*. For now, we'll close the pop-up by clicking on the *X* on the corner.

The second image shows a 1Mb region around our selected region. This is always 1Mb in human, but the fixed size of this view varies between species. This view allows you to scroll back and forth along the chromosome.

Region in detail Ø

nd -
1
ENSG000
5
N 10

You can also drag out and jump to or mark a region.



Click on the *X* to close the pop-up menu.

Click on the *Drag/Select button* **Drag/Select button Drag/Select button Drag/Select button Drag/Select button by** clicking and dragging within the image. As you do this you'll see the image below grey out and two blue buttons appear. Clicking on *Update this image* would jump the lower image to the region central to the scrollable image. We want to go back to where we started, so we'll click on *Reset scrollable image*.

🌣 🔝 < 🕀 🖬 🌜	2									D	rag/Select: ↔ 📋
Chromosome bands	122.87Mb	12	88Mb	Update this image	22.905	Reset scrollable image	122.92Mb		122.93Mb		Forward strand -
Human cDNAs (RefSeg/ENA)								-			

The third image is a detailed, configurable view of the region.



Here you can see various *tracks*, which is what we call a data type that you can plot against the genome. Some tracks, such as the transcripts, can be on the forward or reverse strand. Forward stranded features are shown above the blue contig track that runs across the middle of the image, with reverse stranded features below the contig. Other tracks, such as variants, regulatory features or conserved regions, refer to both strands of the genome, and these are shown by default at the very top or very bottom of the view.

You can use click and drag to either navigate around the region or highlight regions of interest, Click on the *Drag/Select* option at the top or bottom right to switch mouse action. On *Drag*, you can click and drag left or right to move along the genome, the page will reload when you drop the mouse button. On Select you can drag out a box to highlight or zoom in on a region of interest.

With the tool set to *Select*, drag out a box around an exon and choose *Mark region*.

🌣 📑 < 🕀 🖬 🏷	7
R emaneshiahli	127.82Mb 01 here
Human cDNAs (RefSeg/ENA)	and here
	Showing all 16 feature click to show
EST cluster (Unigene)	
	Highlight
	Showing all 30 features - click to shov
CCDS set	CCDS34059.1 >

The highlight will remain in place if you zoom in and out or move around the region. This allows you to keep track of regions or features of interest.

We can edit what we see on this page by clicking on the blue Configure this page menu at the left.

Configure this page

This will open a menu that allows you to change the image.

Configure Region Image Configure Overvie	ew Image	Configure Chromone Image Personal Data	
Find a track			
Active tracks		Select from available onfigurations:	
Favourite tracks		Configuration tabs	
Track order Centrola		Active thacks if gur account i was	
Search results		Sequence and assembly	
Genome Reference Consortium Issues	(0/17)	PI Contine	+ 0
E Sequence and assembly	(3/18)	M Sequence	× •
 Sequence Markers 	(2/4) (0/1)	Primary assembly manning	× 0
- GRC alignments	(1/2)		× U
 Simple features 	(0/4)	Genes and transcripts	
Clones & misc. regions	(0/7)	Comprehensive Gene Annotations from GENCODE 38	0 🦟
Genes and transcripts Genes	(2/77)	CCDS set	* 0
- Long reads	(0/14)	mBNA and protein alignments	
 Prediction transcripts 	(0/1)	Track informa	tion
E LRG BNASeg models	(0/1)	EST Cluster (Unigene)	
E mBNA and protein alignments	(2/7)	In Human CDINAS (Helbergreina)	× 0
- mRNA alignments	(2/3)	Variation	
 EST alignments 	(0/1)	🖬 gnomAD - short variants (SNPs and indels)	* 0
 Protein alignments 	(0/3)	All phenotype-associated - short variants (SNPs and indels)	* 0
E Variation	(3/84)	1000 Genomes 3 - All (structural variants)	* 0
 Phenotype, disease and curated variants 	(1/20)	Regulation Trooph to propo	
 Arrays and other 	(0/17)	IT CER MUMES	
 Failed variants Structural variants 	(0/1)	aliatory Build	* 9
	(0/5)	Comparative genomics	
- Somatic variants	(0/2)	Constrained elements for 90 eunerian mammal EP 202 stended	* 0
 Somatic structural variance 	(0/3)	when Tracks on/off and change style	* 0
Regulation	(1/467)	Information and descerations	
- Regulatory Bur - Features by Calking c b cot	eon		
 DNA methylation 	(947)		* 0
Other regulations regions	10145	In Chromosome bands	EA

There are thousands of possible tracks that you can add. When you launch the view, you will see all the tracks that are currently turned on with their names on the left and an info icon on the right, which you can click on to expand the description of the track. Turn them on or off, or change the track style by clicking on the box next to the name. More details about the different track styles are in this FAQ: http://www.ensembl.org/Help/Faq?id=335.

You can find more tracks to add by either exploring the categories on the left, or using the *Find a track* option at the top left. Type in a word or phrase to find tracks with it in the track name or description.

Let's add some tracks to this image. Add:

• Proteins (mammal) from UniProt - Labels

1000 Genomes - All - short variants (SNPs and indels) – Normal

Now click on the tick in the top left hand to save and close the menu. Alternatively, click anywhere outside of the menu. We can now see the tracks in the image. The proteins track is stranded, so you will see two tracks, one above and one below the contig, representing the proteins mapped to the forward and reverse strands respectively. The variants track is not stranded, so is found near the bottom of the image.

If the track is not giving you can information you need, you can easily change the way the tracks appear by hovering over the track name then the cog wheel to open a menu. To make it easier to compare information between tracks, such as spotting overlaps, you can move tracks around by clicking and dragging on the bar to the left of the track name.

	retained intr	on
All phenotype-assoc	All phenotype-associated - short variants (SNPs and indels)	0
CCDS set	0 \$ 8 * X J	
EST_cluster_(Unigene)	Change track atyle:	ł
	Normal (collapsed for windows over 200kb) Collapsed	
	Expanded with name (hidden for windows over 10kb)	- 1
Human cDNAs (RefSeg/ENA)	Expanded without name	_

Now that you've got the view how you want it, you might like to show something you've found to a colleague or collaborator. Click on the *Share this page* button to generate a link. Email the link to someone else, so that they can see the same view as you, including all the tracks you've added. These links contain the Ensembl release number, so if a new release or even assembly comes out, your link will just take you to the archive site for the release it was made on.

Share this page

To return this to the default view, go to *Configure this page* and select *Reset configuration* at the bottom of the menu.

Human genes and transcripts in Ensembl, Demo

You can find out lots of information about Ensembl genes and transcripts using the browser. If you're already looking at a region view, you can click on any transcript and a pop-up menu will appear, allowing you to jump directly to that gene or transcript.



Alternatively, you can find a gene by searching for it. You can search for gene names or identifiers, and also phenotypes or functions that might be associated with the genes.

We're going to look at the human *UQCRQ* gene. From <u>ensembl.org</u>, type *UQCRQ* into the search bar and click the *Go* button. You will get a list of hits with the human gene at the top.

Where you search for something without specifying the species, or where the ID is not restricted to a single species, the most popular species will appear first, in this case, human, mouse and zebrafish appear first. You can restrict your query to species or features of interest using the options on the left.



The gene tab

Click on the gene name or Ensembl ID. The Gene tab should open:

Human (GRCh38.p13)	• <u> </u>				
Location: 5:132,866,630-132,868,847	Gene: UQCRQ				
Gene-based displays		Gene tab			
Summary	Gene: UQCRQ EN	NSG00000164405			
 Splice variants 					
 Transcript comparison 	Description	ubiquinol-cytochrome c reductase complex III subunit VII [Source:HGNC Symbol;Acc:]			
- Gene alleles	Gene Synonyms	OCB8 OP-C UQCB7			
Secondary Structure	Lono ognolignio				
□ Comparative Genomics	Location	<u>Chromosome 5: 132,866,630-132,868,847</u> forward strand.			
 Genomic alignments 		GRCh38:CM000667.2			
- Gene tree	About this gene	This gene has 6 transcripts (splice variants), 209 orthologues and is associated with 3			
 Gene gain/loss tree 	Transcripts				
- Orthologues		Show transcript table			
- Paralogues					
	Summary @	Option open table of tra			
GO: Molecular function		option open table of the			
- GO: Cellular component	Name	UQCRQ 굗 (HGNC Symbol)			
- GO: Biological process	CCDS	This gene is a member of the Human CCDS set: CCDS34237.1			
- Genetic Variation	UniProtKB	This gene has proteins that correspond to the following UniProtKB identifiers: 014940			
- Variant table	D. (O	The generate process and correspond to the following offic for the full littles.			
- Variant image	RefSeq	This Ensembl/Gencode gene contains transcript(s) for which we have selected identic			
Structural variants Dene	wews.	they will be in the <u>External references</u> table			
- Gene expression	Ensembl version	ENSG00000164405.11			
- Regulation	Other assemblies	This gene maps to 132,202,322-132,204,539 in GRCh37 coordinates.			
- External references		View this locus in the GRCh37 archive: ENSG00000164405			
 Supporting evidence 	Gene type	Protein coding			
ID History	sene type				
└─ Gene history	Annotation method	Annotation for this gene includes both automatic annotation from Ensembl and Havan			
Configure this page	Go to Regi	ion in Detail for more tracks and navigation options (e.g. zooming)			
	★ 24 ~ 世 単 ☆ `	8			
Share this page		22.22 kb			
R Bookmark this page	Genes (Comprehensive set	Transcripts: click			
	Contigs	for more info UQCRQ-203 > protein coding UQCRQ-201 > protein coding UQCRQ-201 > retained intron < AC004500.1			

This page summarises the gene, including its location, name and equivalents in other databases. At the bottom of the page, a graphic shows a region view with the transcripts. We can see exons shown as blocks with introns as lines linking them together. Coding exons are filled, whereas non-coding exons are empty. We can also see the overlapping and neighbouring genes and other genomic features.

There are different tabs for different types of features, such as genes, transcripts or variants. These appear side-by-side across the blue bar, allowing you to jump back and forth between features of interest. Each tab has its own navigation column down the left hand side of the page, listing all the things you can see for this feature.

Let's walk through this menu for the gene tab. How can we view the genomic sequence? Click *Sequence* at the left of the page.



The sequence is shown in FASTA format. The FASTA header contains the genome assembly, chromosome, coordinates and strand (1 or -1) – this gene is on the positive strand.

	Marked-up sequence M BLAST this sequence	AST or download is sequence
	Exons UQCRQ exons All exons in this region Markup loaded	Upstream sequennce
	>chromosome:GRCh38:5:132866030:132869447:1 TGTATTATCTTGGGAACGTTAACAGCCTGCCAGGGGACTCTAAAACAGAGTCC AAGCCTTCGATGTGTCCCTGATTTGGTGCTGGGGTCGCTCATGGCCTGAAAA AAGCCCCCCTTTGAGCTACTGGCCAAAGGCCAAATGAAAC <mark>CTCGTGCTCCTCTGCTG CGGCCCGGCCTCGGGGCTCAGAGGCCCTCTGTCTACACCTGGGAACACTCGCAGG</mark>	AAGAC TAACG TGGGG CCCCA
Exon of an (AGCTCCTGGCCAGCGCTGCCCGGAGGTCGGCAGGCCCCTTCCTCGTCACCTTTT CCTCCCCGCCTCCCGCATTCGGCCGCTTCCTGACTGGGATTCCACAGAAAAGCC CTGAGGAGAAGTGTGAGCGCCTCCGCCTGTCCACTGTCCCCCAAAGTCAGTTCAA	TGTTC GAGGG TCCCC
overlapping	GACGTCCTCCGCTAGGCTCCACCCCACCGGCCCGGGCAGGGCCTCCAAGGCACCT CTACGGTCACCCAGTCAGCCCACTTCTTCTGGGACAAAGGCGTCATCCCTTAG GTAGGAAATGGTATCTCCCGGAGGTCACCTCCACGACGTCGCGGCGAGCGA	agaca UQCRQ exon
gene	gcgtgagtggggtttggttgtgcagtgttcgggccctgggaggctaggggcgc tgggctgggaaaggataaggagtgcaggggaggctggggtggggatggggatggg	

Exons are highlighted within the genomic sequence, both exons of our gene of interest and any neighbouring or overlapping gene. By default, 600 bases are shown up and downstream of the gene. We can make changes to how this sequence appears with the blue *Configure this page* button found at the left. This allows us to change the flanking regions, add variants, add line numbering and more. Click on it now.

Configure Page Personal Data			
Display options	Select from available configurations:	Current unsaved ~	Save current configuration
Reset configuration	Display options		
	5' Flanking sequence (upstream):	600	* (Maximum of 1000000)
	3' Flanking sequence (downstream):	600	* (Maximum of 1000000)
	Number of base pairs per row:	60 bps 🗸	
	Additional exons to display:	Core exons 🗸	
	Orientation of additional exons:	Display exons in both orientation ~	
	Show variants:	No	5
	Hide variants longer than 10bp:		Show variants
	Hide variants by frequency (MAF):	Don't hide 🗸	
	Filter variants by consequence type:	No filter 3 prime UTR variant 5 prime UTR variant NMD transcript variant coding sequence variant	Eilten to only 10
	Filter variants by evidence status:	No filter 1000Genomes Cited ESP ExAC	Pliter to only 10
	Hide individual variant sources:	No filter Hide AMDGC Hide Affy GenomeWideSNP_6 C Hide Archive dbSNP	Show line numbe
	Line numbering:	None ~	

Once you have selected changes (in this example, *Show variants*, *1000 Genomes variants* and *Line numbering*) click at the top right.

	Liv	nks to variant tab
421	GACGTCCTCC	480 <u>431: rs563868690</u> <u>450: rs15</u> 867588
481	CTACGGGTCACCCAGTCAGCCCACTTCTTTCTGGGACAAAGGCGTCATCCCT Y AGAGAC <mark>N</mark>	540 <u>533: rs114932352</u> <u>540: rs60321536</u>
541	DTAGGAAAATRGTATCTCCCGGAAGTTACCTCACGACCTCCAAGAGCNGCTTCCAACCTT	600 <u>541: rs192032018</u> <u>551: rs548641026</u>
601	GCCGGAAATGACGAACGAGTCAACCGGATCGGTGACTGTGGAGGGCGAGCTGA <mark>N</mark> C <mark>I</mark> CTGT	660 <u>654: rs527907338</u> <u>656: rs369119794</u>

You can download this sequence by clicking in the Download sequence button above the sequence:

🛃 Download sequence

This will open a dialogue box that allows you to pick between plain FASTA sequence, or sequence in RTF, which includes all the coloured annotations and can be opened in a word processor. If you want run a sequence analysis tool, download as FASTA sequence, whereas if you want to analyse the sequence visually, RTF is best for this. This button is available for all sequence views.

Download sequence		
File name:	Homo_sapiens_UQCRQ_sequence	
File format:	FASTA	
	O Preview 🛃 Download	Download Compressed
Settings		
Sequences to export:	 Select/deselect all cDNA (transcripts) Coding sequences (CDS) Amino acid sequences 5' UTRs 3' UTRs Exons Introns Genomic sequence 	
5' Flanking sequence (upstream):	600	* (Maximum of 1000000)
3' Flanking sequence (downstream):	600	* (Maximum of 1000000)
Fields marked * are required Guide to file formats		
TASTA HIF Text sequence(s): Market DNA and/or amino acids *11 dna:chrowosome chromosome (atch8:11:10) *11 dna:chrowosome chromosome (atch8:11:10) ATTAC *12 dna:chrowosome chromosome (atch8:11:10) ATTAC *13 dna:chrowosome chromosome (atch8:11:10) ATTAC *14 dna:chrowosome chromosome (atch8:11:10) ATTAC *15 dna:chrowosome chromosome (atch8:11:10) ATTAC *16 dna:chrowosome chromosome (atch8:11:10) ATTAC *17 dna:chromosome chromosome (atch8:11:10) ATTAC *17 dna:chromosome chromosome (atch8:11:10) ATTAC *18 dna:chromosome chromosome (atch8:11:10) ATTAC *17 dna:chromosome chromosome (atch8:10) ATTAC *18 dna:chromosome chromosome (atch8:11:10) ATTAC *11 dna:chromosome chromosome (atch8:11:10) ATTAC *11 dna:chromosome chromosome chromosome (atch8:11:10) ATTAC *11 dna:chromosome chromosome (atch8:11:10) AGGCC <	ed-up sequence, pr without variants CAACAAAAAGCAAACACGGG TCTTCCACAAACATGGGCAT GGGTAATGTGGCTTTCCGT GGGTAATGTGGCTTTCCGT CTCACAATTCGTCCAAGTG GTTTCCGCACCTGGGACCTC GGCATGTGGAGCTGATGCTT	

To find out what the protein does, have a look at GO terms from the <u>Gene Ontology consortium</u>. There are three pages of GO terms, representing the three divisions in GO: Biological process (what the protein does), Cellular component (where the protein is) and Molecular function (how it does it). Click on *GO: Biological process* to see an example of the GO pages.

GO: Biological process @

Find other genes

			•	
Show/hide colu	mns (1 hidden)		linked to this te	rm 🔨 🕺
Accession	🔶 Term 🔶	Evidence Annotation source	Transcript IDs	÷ 🔰
<u>GO:0006122</u> 교	mitochondrial electron transport, ubiquinol cytochrome c	BA GO_Central	ENST00000378667 ENST00000378670 ENST00000378665	 <u>Search BioMart</u> <u>View on karyoty</u>
<u>GO:0021539</u> ਲੋ	Hourermoueroptine codes	EGO termsoare link	to <u>ENST00000378670</u>	Search BioMart View on karyoty
<u>GO:0021548</u> &	forsdefinitions	Etranscripts, not g	genes ===================================	 Search BioMart View on karyoty
<u>GO:0021680</u> &	cerebellar Purkinje cell layer development	See which ones	ENST00000378670	 <u>Search BioMart</u> View on karvoty

Here you can see the functions that have been associated with the gene. There are three-letter codes that indicate how the association was made, as well as links to the specific transcript they are linked to.

We also have links out to other databases which have information about our genes and may focus on other topics that we don't cover, like Gene Expression Atlas or OMIM. Go up the left-hand menu to *External references*:

External references @

Download all tables as CSV

This gene corresponds to the following database identifiers:

	Filter 🖬
External database	Database identifier
Expression Atlas	ENSG0000164405 & [view all locations]
HGNC Symbol	UQCRQ 군 ubiquinol-cytochrome c reductase complex III subunit VII [view all locations]
MIM gene	UBIQUINOL-CYTOCHROME c REDUCTASE, COMPLEX III SUBUNIT VII, 9.5-KD; UQCRQ [*612080] @ UBIQUINOL-CYTOCHROME c REDUCTASE, COMPLEX III SUBUNIT VII, 9.5-KD; UQCRQ;;QPC [view all locations]
MIM morbid	MITOCHONDRIAL COMPLEX III DEFICIENCY, NUCLEAR TYPE 4; MC3DN4 (#615159) & MITOCHONDRIAL COMPLEX III DEFICIENCY, NUCLEAR TYPE 4; MC3DN4 [view all locations]
NCBI gene (formerly Entrezgene)	UQCRQ 문 ubiquinol-cytochrome c reductase complex III subunit VII [view all locations]
Reactome gene	R-HSA-1428517 The citric acid (TCA) cycle and respiratory electron transport [view all locations] R-HSA-1430728 Metabolism [view all locations] R-HSA-163200 Respiratory electron transport, ATP synthesis by chemiosmotic coupling, and heat production by uncoupling proteins. [view all locations] R-HSA-611105 Respiratory electron transport [view all locations]
WikiGene	UQCRQ 문 ubiquinol-cytochrome c reductase complex III subunit VII [view all locations]

Demo: The transcript tab

We're now going to explore the different transcripts of UQCRQ. Click on Show transcript table at the top.

Show transcript	Show transcript table											
Show/hide columns (1 hidden)												
Transcript ID 💧	Name 💧	bp 🗄	Protein 🖕	Biotype	CCDS 🖕	UniProt Match 🖕	RefSeq Match 🖕	Flags				
ENST00000378670.8	UQCRQ-203	1573	<u>82aa</u>	Protein coding	CCDS34237	<u>014949</u>	NM_014402.5	MANE Select v0.95 Ensembl Canonical GENCODE basic APPRIS P1				
ENST00000378665.1	UQCRQ-201	586	<u>82aa</u>	Protein coding	CCDS34237	<u>014949</u> &		GENCODE basic APPRIS P1 TSL:1				
ENST00000378667.1	UQCRQ-202	429	<u>82aa</u>	Protein coding	CCDS34237	<u>014949</u>		GENCODE basic APPRIS P1 TSL:2				
ENST00000496429.1	UQCRQ-205	258	No protein	Processed transcript		-		TSL:2				
ENST00000498309.1	UQCRQ-206	874	No protein	Retained intron	-	-		TSL:2				
ENST00000480372.1	UQCRQ-204	509	No protein	Retained intron	-	-	-	TSL:2				

Here we can see a list of all the transcripts of *UQCRQ* with their identifiers, lengths, biotypes and flags to help you decide which ones to look at.

If we were to only choose one transcript to analyse, we would choose UQCRQ-203 because it is the MANE Select and Ensembl Canonical. This means it is both 100% identical to the RefSeq transcript NM_014402.5 and both Ensembl and NCBI agree that it is the most biologically important transcript.

Click on the ID, ENST00000378670.8.

You are now in the *Transcript tab* for UQCRQ-203. We can still see the gene tab so we can easily jump back. The left hand navigation column provides several options for the transcript UQCRQ-203 - many of these are similar to the options you see in the gene tab, but not all of them. If you can't find the thing you're looking for, often the solution is to switch tabs.

Click on the *Exons* link. This page is useful for designing RT-PCR primers because you can see the sequences of the different exons and their lengths.

C	Show/hide columns)					Filter
No.	Exon / Intron	Start	End	Start Phase	End Phase	Length	Sequence
	5' upstream sequence						aagttacctcacgacctccaaga agettccaaccttgccggaaatgac
1	ENSE00001813218	132,866,642	132,866,687	-	-	46	GAACGAGTCAACCGGATCGGTGACTGTGGACGGCGAGCTGAGCCCT
	Intron 1-2	132,866,688	132,866,868			181	gtgcgtgagtggggtctgoggech: flankingcsequenceag
2	ENSE00001084104	<u>132,866,869</u>	132,867,035	ran	ge: (UTR	GGCCGCCGCCACAATGGGCCGCGAGTTTGGGAATCTGACGCGGATGCGGCATGTGATCAG CTACAGCTTGTCACCGTTCGAGCAGCGCGCCTAC GCACGTCTTCACTAAAGGAATCCC CAATGTTCTGCGCCGCATTCGGGAGTCTTTCTTTCC GGGGGGCCCGC
	Intron 2-3	132,867,036	132,867,487			452	gtgagtgccttgggcccgcgggggggggggggggggggg
3	ENSE00001478312	<u>132,867,488</u>	<u>132,868,847</u> 9	1 rey:	int	1, 60 rons	AGTTTGTAGTGTTTTATCTTATCTACACATGGGGGACTGAGAGATCCAAGA GGAAGAATCCAGCTGCCTATGAAAATGACAAATGAGCAACGCATCCGGATGACGGTTCCC TGTCTCTGAAAGACCTTTCTCTGGAAGAGGAGTCTGCATTGTAGTGTCTCCAAAGACACAA

You may want to change the display (for example, to show more flanking sequence, or to show full introns). In order to do so click on *Configure this page* and change the display options accordingly.

Now click on the *cDNA link* to see the spliced transcript sequence with the amino acid sequence. This page is useful for mapping between the RNA and protein sequences, particularly genetic variants.

	KY K RRRR Y H* S YR *SW **R Y R S **Y*H***MYY** M	
181	<mark>gCat</mark> tcg <mark>ggagtc</mark> ttt <mark>c</mark> ttt <mark>c</mark> tt <mark>cgcg</mark> tg <mark>gt</mark> gccg <mark>cag</mark> tt <mark>tgtagtgtttt<mark>tatcttatcta</mark>ca</mark>	240
122	GCATTCGGGAGTCTTTCTTTCGCGTGGTGCCGCAGTTTGTAGTGTTTTATCTTATCTACA	181
41	RIRESFFRVVPQFVVFYLIY	60
	<u>Y RR S R * *R * RR S R Y Y R R BV</u>	
241	CA <mark>T</mark> GG <mark>GG</mark> GA <mark>CTG</mark> AAGAGTT <mark>CG</mark> AGAGATCCA <mark>AG</mark> AGGAAGAAT <mark>C</mark> C <mark>A</mark> GCTG <mark>C</mark> CT <mark>A</mark> TGA <mark>A</mark> AA <mark>TG</mark>	300
182	CATGGGGGACTGAAGAGTTCGAGAGATCCAAGAGGAAGAATCCAGCTGCCTATGAAAATG	241
61	TWGTEEEESKRKNPAAYEN	80

UnTranslated Regions (UTRs) are highlighted in dark yellow, codons are highlighted in light yellow, and exon sequence is shown in black or blue letters to show exon divides. Sequence variants are represented by highlighted nucleotides and clickable IUPAC codes are above the sequence.

Next, follow the *General identifiers* link at the left. Just like the *External References* page in the gene tab, this page shows links out to other databases such as RefSeq, UniProtKB, PDBe and others, this time linked to the transcript or protein product, rather than the gene.

General identifiers @

This transcript corresponds to the following database identifiers:

Show All 🗸 entries		Filter	
External database	Database identifier		
BioGRID Interaction data, The General Repository for Interaction Datasets	117991 [view all locations]		
CCDS	CCDS34237.1 & [view all locations]		
European Nucleotide Archive	BC001390 & [align] [view all locations] BC090048 & [align] [view all locations] D50369 & [align] [view all locations]		
Human Protein Atlas	ENSG00000164405 준 [view all locations] HPA046693 준 [view all locations] HPA053323 준 [view all locations]		
INSDC protein ID	AAH01390.1 & [align] [view all locations] AAH90048.1 & [align] [view all locations] BAA23321.1 & [align] [view all locations]		

If you're interested in protein domains, you could click on *Protein summary* to view domains from Pfam, PROSITE, Superfamily, InterPro, and more. These are all plotted against the transcript sequence, with the exons shown in alternating shades of purple at the top of the page. Alternatively, you can go to *Domains & features* to see a table of the same information.

Protein summary @

Superfamily Pfam

PANTHER

Alternating shades of purple Protein domains for ENSP00000367939.3 show the exon structure 🔅 < 🖬 🇞 🏷 ENSP00000367939.3 PDB-ENSP mappings chrome b-c1 complex sub Durtain 1----

-	Gene3D	Cytoc	hrome b-c1 con	mplex subunit	8 superfamily		Protein	domai	ns			-
	All sequence SNPs/i	Sequenc KKKR RMY S	ce variants (dl MYSHRYY YVSH V S	bSNP and all MVYSKY YV D YD D D	other sources) SY REYRRY EKYB R NB N N	KMYMWKBK SV MBY V S Y	(RRYH SRW <mark>R</mark> R <mark>ih y</mark> S R	YR SPYM SM YY	YRSR R	R	SYYRRBV R BVI Y	2
	Variant - COSMIC s	Variant ·	- COSMIC son	natic mutatio	ons					I	•	
	Variant Legend	stop stop infr mis syn	p gained p lost ame insertion sense varian onymous var	n t riant			frameshift v start lost inframe dele splice region coding sequi	ariant tion variant ence variant				-
	Scale bar	0	8	16	24	32 40	48	56	64	7	2 82	2

You can also see the structure of the protein from the PDB by clicking on PDB 3D Protein model.

3D Protein model (PDBe) @

Select PDBe model: 5xte - Coverage: [PDBe: 2-82 | ENSP: 2-82] => 99% of ENSP length V 3D Viewer Help 2 0 0 \$ ENSP00000367939 -Ensembl-PDBe mapp Label **Q** PDB Coverage 2-82 Exons Exons Protein information Gene3D PANTHER Pfam Variants SIFT PolyPhen n<mark>t exons</mark>, variants **_** d domaⁱns

This uses LiteMol to show a 3D protein. You can use all the normal controls that you would use with LiteMol, plus plot Ensembl features like Exons and variants onto the structure using the options on the right. We allow you to see the top ten PDB models for this protein, based on coverage and quality scores, you can choose which at the top of the viewer.

Ensembl protein: ENSP00000367939 | PDBe model: 5XTE @ (Cytochrome b-c1 complex subunit 8 - chains A,N)

Exploring variants in Ensembl, Demo

In any of the sequence views shown in the Gene and Transcript tabs, you can view variants on the sequence. You can do this by clicking on Configure this page from any of these views.

Let's take a look at the Gene sequence view for *HBB* in human. Search for *HBB* and go to the *Sequence* view.

If you can't see variants marked on this view, click on *Configure this page* and select *Show variants: Yes and show links*. You may also wish to add a filter to the variants to allow them to load more quickly, we'll add *Filter variants by evidence status: 1000Genomes*.

Marked-up sequence @	
Download sequence	
Exons DHDDS exons All exons in this region	
Variants 3 prime UTR 5 prime UTR Intronic Missense Non-coding exon	Regulatory region Splice region
Synanymous Tf binding site	
Markup landad	
maikup loaded	
Variant loogu	1
Filters applied	4
Only showing variants with evidence status: 1000Genomes	
>chromosome:GRCh38:1:26431682:26471906:1	
1 CTAGAAGTCATTCTTATGGCAGAARTCTTTCTGGCCTTACCCCTTCTCTTGAGTTCGTTG	60 <u>25: rs560479343</u>
61 GGATAGGTCTTGGGAATCCACAAATCCALA ATAGCGTGCCTTAAGTCGGCACAAAATA 121 GATEBATATTCTCTGGGCCTCTCTBAGATTGCCCCCCCTAGBBCTCCTTCCCCCGCCCATT	120
181 ANATCCACTAMATANAATACGAAACGGTAACC GATAATTTTTTAAGAGTATGCACCAGTG	240 191: rs546488329
241 CTCTAMGAAGCCCGGTTCTGCGCCATTCCAAGGTAAGGACGCTGCCTTACCCAGTCTCT	300 247: rs148143529
301 CAGCAGTCCTAAGCTGC TICLEO, CLACAAGCGCCTCLAR, CTTGCCOTGGCTCAT	360 341: 0053 232627
361 TTCCCTGGCGCGTTTT CACAGE CONTENTS AND A GALGE UCTTCCC	420 3 9 1 1 N S 09 01: rs191286307
421 CAGCTGGTCNCCCAGCACGTEGCCTGGAGCCCAGCCTCCCGCCCGCTGGGTCACGTGAG	480 430: rs3811462 441: rs141969222 46
481 CCTGGCTCTCACCCGCGCCGGCTTCCACACCCCCCGGGACTGAGAACCGCCCTCCCACTA	540 <u>514: rs538561781</u>
541 GCTCGCAGGCGCAGGCGCAGGGCAGGGTCCCCTCAGTTTGGGTRACGGAAGGGGCGTGTC	600 <u>563: rs115743942 576: rs3828083</u> <u>58</u>
601 PCAGCGYCCAAAGATGCCOCCACT	660 ναγια[Ιση Ταθ 62
661 GRAGCCGTGGGGACGCGCCCAGCGGAGCTAATCAGGTGCCACTGAACTGAAGGGGTGAAC	720
721 TACAGGAGAAGAAGAGGGTGACTGGGCGGGGAGCAGCTGCGGGAGAAGCAAAGGGACGACTGA	780 <u>723: rs561032073</u>
841 BGBARTARTURGGACCACTGROOC TGRAGTACTAGGCGTGCGATACTGRAGGAGTT	040 001: 18140190301
0.11 November 10.000 11.000 11.000 0001 1000 0001 0000 0000 0000 0000 0000 0000 0000 0000	300

Find out more about a variant by clicking on it.



You can add variants to all other sequence views in the same way.

You can go to the Variation tab by clicking on the *variant ID*. For now, we'll explore more ways of finding variants.

To view all the sequence variants in table form, click the Variant table link at the left of the gene tab.

Variant table Ø

Variant table																	0
This table sho	This table shows known variants for this gene. Use the 'Consequence Type' filter to view a subset of these.																
Filter 🝸 Giobal MAF: All 🝸 SIFT: All 🍸 PolyPhen: All 💥 Consequences: missense 🐂 T																	
							Show/h	ide celumps				4.			Search		
Variant ID	Chr: bp	Alle- les	Glo- bal MAF	Class	Sour-	Evid- ence	Clin. Sig.	Conseq. Type	AA	AA co- ord	SIFT	Poly- Phen	CAD	REV	Meta	Muta tion Asse ssor	Transcript
rs33985739	11:5225601	G/ C /T	-	SNP	dbSNP	5	?	missense variant	H/Q	147	0.03	0.351	22	0.828	0.788	9	ENST000003: 5295.4
rs33985739	11:5225601	G/C/T		SNP	OSNP	5%	?	missense variant	H/Q	147	0.03	0.351	22		0.788	0.879	ENST000003
rs33954264	11:5225602	T/ A /C/ G	-	Evide	hee	code	? 2?	missense variant	H/L	147	0	0.76	24	0.865	0.874	0.955	UNS10000033 5295.4
rs33954264	11 5225602	T/A/ C / G		SNP	dbSNP	5%	▲?	missense variant	H/R	147	0.01	0.76	24	a	fec	ted	ENST0000033 5295.4
153395 Jal	-iant	T//C	S	SNP	dbSNP	5	▲?	Hat	ho	gen	icit	JY.975	CO	res	0.908	0.955	ENST000003

You can filter the table to only show the variants you're interested in. For example, click on *Consequences: All*, then select the variant consequences you're interested in. For display purposes, the table above has already been filtered to only show missense variants.



You can also filter by the different pathogenicity scores and MAF, or click on *Filter other columns* for filtering by other columns such as Evidence or Class.

The table contains lots of information about the variants. You can click on the IDs here to go to the Variation tab too.

You can also see the phenotypes associated with a gene. Click on *Phenotype* in the left hand menu.

Phenotypes @

Phenotypes, diseases and traits associated with this gene ENSG00000244734

Show 10 v entries	Filter
Phenotype, disease and trait	Source
ALPHA-THALASSEMIA	MIM morbid @
BETA-THALASSEMIA	MIM morbid #
BETA-THALASSEMIA INTERMEDIA	Orphanet &
Beta-thalassemia major	Orphanet @
BETA-THALASSEMIA, DOMINANT INCLUSION BODY TYPE	DI MIM morbid tS
Delta-beta thalassemia	Phenolypes linked
Dominant beta-thalassemia	Orphanet &
ERYTHROCYTOSIS, FAMILIAL, 6	MIM morbid #
FETAL HEMOGLOBIN QUANTITATIVE TRAIT LOCUS 1	To the gene MIM morbid #
HEINZ BODY ANEMIAS	MIM morbid #
Showing 1 to 10 of 26 entries	< < 1 2 3 > 20

Phenotype, disease and trait annotations associated with variants in this gene

Show 10 🗸 entries				Filter
Phenotype, disease and trait	Source(s)	Number of v	ariants 🖕	Show/hide details
ALL variants with a phenotype annotation (WARNING: details table may not load for this number of variants!)	· · ·	F .	1355	Show
Anemia	ClinVar d?		2	Show
Annotated by HGMD	HGMD-PUBLIC		715	Show
BETA-PLUS-THALASSEMIA	ClinVar @	D/ A	36	Show
BETA-PLUS-THALASSEMIA, DOMINANT	ClinVar #	Phenotype	s lin	Redition
BETA-THALASSEMIA INTERMEDIA, DOMINANT	ClinVar @		1	Show
BETA-THALASSEMIA, DOMINANT INCLUSION BODY TYPE	ClinVar #		16	Show
Beta thalassemia intermedia	ClinVar 🗗	to variant	sin	Show
Beta thalassemia major	ClinVar #	•	3	Show
Beta-houston-thalassemia	ClinVar P		1	Show
Showing 1 to 10 of 466 entries		the gene		2 3 4 5 > >>

Phenotype, disease and trait annotations associated orthologues of this gene in other species

Show All - entries			DI A Filter In / 🎘
Phenotype, disease and trait	Source	Species	Khenotypes linked
anatomical system, quality	<u>ZFIN</u> ₽	Zebrafish (Danio rerio)	ENSDARG00000088530 hbse1.3
liver, quality	ZEIN@	Zebrafish (Danio rerio)	to orthologues
intestine.quality	ZEIN®	Zebrafish (Danio renio)	ENSDARG00000088330 https://a
whole organism, shape	ZFIN @	Zebrafish (Danio rerio)	ofestheogene

Open the transcript table and go to HBB-201 ENST00000335295, then click on *Haplotypes* in the left hand menu.

Haplotypes @

Export data as JSON	Switch to	CDS view							
Show All 🗸 entries		Show/	hide columns	01	20		Filter		
Protein haplotype	 Flags 	- General	protein	onCL	LAS	EUR	SAS	GGVP:ALL	÷
REF		0.952 (5727)	0.887 (1172)	0.99 (687)	0.985 (993)	0.998 (1004)	0.989 (967)	0.895 (904)	
<u>7E>V</u>	D	0.0382 (230)	0.0998 (132)	0.0072 (5)	0 (0)	0 (0)	0(0)	0.0921 (93)	
7E>K		0.00449 (27)	0.0129 (17)	0 (0)	0 (0)	0.0	0(0)	0.0099 (10)	
27E>K		0,00233 (14)	Q (0)	0 (0)	0.00794 (8)	0 (0)	0.00613 (6)	0 (0)	
18K>*.19del(130)	• • (i	kelveto al	ectorot	'ein	0.00595 (6) 🖊	romun	819 :10	1000	
122E>Q	Cit	0.000499 (3)	0 (0)	0 (0)	0 (0)	requer	0,0307 (3)	0 (0)	
Variant	's that	t are foun	d		6	enome	s popi	ulation	ns

together in individuals

The Haplotypes view in the transcript tab shows you the actual protein and CDS sequences in 1000 Genomes individuals. This is possible because the 1000 Genomes study has phased genotypes, so we know which alleles occur on which of the chromosome pairs. The table lists all the versions of the protein that occur along with their frequencies, including the reference sequence and sequences with one or more alternative alleles.

Click on one of the haplotypes, we'll go for *18K>*,19del{130}*, to find out more about it. Here you will see the frequency in the 1000 Genomes subpopulations, the sequence and the 1000 Genomes individuals where this protein is found.

Let's have a look at variants in the Location tab. Click on the Location tab in the top bar.

Configure this page and open Variation from the left-hand menu.

Configure Region Image Configure Overv	iew Image	Configure Chromosome Image Personal Data							
Find a track		Select from available configurations:	Current unequal Y Grup support configuration						
Active tracks		Select non available configurations.	Current unsaveu						
Favourite tracks		Variation							
Track order		variation							
Search results		Enable/disable all Sequence variants							
Genome Reference Consortium Issues	(0/17)	Sequence variants (dbSNP and all other sources)	equence variants (dbSNP and all other sources)						
Sequence and assembly	(3/18)	dbSNP variants	bSNP variants						
- Markers	(0/1)	ExAC - short variants (SNPs and indels)		* 0					
GRC alignments	(1/2)	UniProt - short variants (SNPs and indels)		÷ 0					
Simple features	(0/4)	gnomAD - short variants (SNPs and indels)		* 0					
Ciones & misc. regions	(0/7)	DECIPHER variants	ECIPHER variants						
Genes and transcripts	(2/77)	LOVD variants	LOVD variants						
Long reads	(0/14)	Mastermind variants							
Prediction transcripts	(0/1)	Eachle/diachle all 1000 Canamas							
 - RNASed models 	(0/1)	Enable/disable all 1000 Genomes	nable/disable all 1000 Genomes						
mRNA and protein alignments	(2/7)	1000 Genomes - All - short variants (SNPs and in	idels)	* 0					
mRNA alignments	(2/3)	1000 Genomes - All - common - short variants (Sl	NPs and indels)	* 0					
- EST alignments	(0/1)	1000 Genomes - AFR - short variants (SNPs and	indels)	* 0					
Protein alignments	(0/3)	1000 Genomes - AFR - common - short variants ((SNPs and indels)	* 0					
Variation Sequence variants	(4/84)	1000 Genomes - AMR - short variants (SNPs and	1 indels)	* 0					
 Phenotype, disease and curated variants 	5 (1/20)	1000 Genomes - AMR - common - short variants	(SNPs and indels)	* 0					
Arrays and other	(0/17)	1000 Genomes - EAS - short variants (SNPs and	indels)	* 0					
 Failed variants Structural variants 	(0/1)	1000 Genomes - EAS - common - short variants ((SNPs and indels)	* 0					
Sometic mutations	(0/5)	1000 Genomes - EUR - short variants (SNPs and	l indels)	* 0					
- Somatic variants	(0/2)	1000 Genomes - EUR - common - short variants	(SNPs and indels)	* 0					
Somatic structural variants	(0/3)	1000 Genomes - SAS - short variants (SNPs and indels)							
Regulation	(1/467)	1000 Genomes - SAS - common - short variants ((SNPs and indels)	* 0					
- Regulatory Build	(1/1)	Enable/disable all Phenotype/disease variants	by source						
 DNA methylation 	(0/47)	All ClinVar variant apportations - short variants /Sh	NPe and indale)	+ 8					
Other regulatory regions	(0/4)		the dense instance/	× 0					

There are various options for turning on variants. You can turn on variants by source, by frequency, presence of a phenotype or by individual genome they were isolated from. You can also turn on genotyping chips.

Let's have a look at a specific variant. If we zoomed in we could see the variant rs334 in this region, however it's easier to find if we put *rs334* into the search box. Click through to open the Variation tab.



The icons show you what information is available for this variant. Click on *Genes and regulation*, or follow the link on the left.

Genes and regulation @

Gene and Transcript consequences

Show 10 🗸 entries	3		Show/hide	columns						Filter					
Gene	Transcript (strand)	Allele (Tr. Gibri	lø)equence Typ	Bosition in tran	scRpsition in CDS	Position in pro	teinAA 🕂	Codons		olyPhen	CADD	REVEL	MetaLR +	Mutation As	se∄
ENSG00000244734 HGNC: HBB	ENST00000335295.4 (-) biotype: protein_coding	A mis (T) var	ssense riant	70 (out of 628)	20 (out of 444)	7 (out of 147)	E/V	GAG/GTG	0.01	0.007	13	0.535	0.071	0.939	S
ENSG0000244734 HGNC: HBB	ENST000003352 (-) biotype: protein_coding	C mis (G) var	ssense riant	70 (out of 628)	20 (out of 444)	7 (out of 147)	E/G	GAG/GGG	0.05	0.009	14	0.472	0.529	0.766	S
ENSG0000244734 HGNC: HBB	ENST00000335295.4 (-) biotype: protein_coding	G mis (C) var	ssense riant	70 (out of 628)	20 (out of 444)	7 (out of 147)	E/A	GAG/GCG	0.29	0.006	5	0.503	0.459	0.326	s
ENSG00000244734 HGNC: HBB	ENST000003803152K1Q	nscrip	iant th	renvar	iant f	allsin	E/V	GAG/GTG	0.01	0.007	13	0.535	0.071	0.939	S
ENSG00000244734 HGNC: HBB	ENST00000380315.2 (-) biotype: protein_coding	C mis (G) var	ssense riant	250 (out of 502)	20 (out of 272)	7 (out of 90)	E/G	GAG/GGG	0.07	0.009		0.472	0.529	0.766	s
ENSG00000244734 HGNC: HBB	ENST00000380315.2 (-) biotype: protein_coding	G mis (C) var	ssense riant	250 (out of 502)	20 (out of 272)	7 (out of 90)	E/A	GAG/GCG	0.34	0.006	5	0.503	0.459	0.326	S
ENSG00000244734 HGNC: HBB	ENST00000485743.1 (-) biotype: protein_coding	A mis (T) var	ssense riant	71 (out of 680)	20 (out of 336)	7 (out of 111)	E/V	GAG/GTG	0.01	0.013	(H30	tho	geni	िर्मि	S
ENSG0000244734 HGNC: HBB	ENST00000485743.1 (-) biotype: protein_coding	C mis (G) var	ssense riant	71 (out of 680)	20 (out of 336)	7 (out of 111)	E/G	GAG/GGG	0.04	0.449		0.472	0.529	0.766	S
ENSG00000244734 HGNC: HBB	ENST00000485743.1 (-) biotype: protein_coding	G mis (C) var	ssense riant	71 (out of 680)	20 (out of 336)	7 (out of 111)	E/A	GAG/GCG	0.31	0.369	5	0.503	0.459	0.326	s
ENSG00000244734 HGNC: HBB	ENST00000633227.1 (-) biotype: nonsense_mediated_decay	A mis (T) var NN var	ssense riant ID transcript riant	70 (out of 609)	20 (out of 168)	7 (out of 55)	E/V	GAG/GTG	0	0.054	13	0.535	0.071	0.939	S
Showing 1 to 10 of	15 entries										« < 1	2 > >>			
Gene expression	correlations			_											
Show 10 🗸 entri	35		Show/hide	columns						Filter		<u></u>			
Gene	\$	P-value (-log ₁	<u>o)</u>		Effect s	ize	•	Tissue					٠.		
ENSG00000132109	•	3.44048238027	790916		0.66455	i3		Th2_memo	ry						
ENSG00000132109		3.22854040129	949654		0.55035	6		Treg_memo	ory						
ENSG00000132109		3.0210752900	19598		0.55224	3		CD8_T-cell	_naive						
ENSG00000132109	Genes up	7.80/8660057	reau	lated	by the	's vari	ant	CD4_T-cell	_naive						
ENSG00000132109	Contra off	2.6244296550	7921		0.53048	14		Treg_naive							
ENSG00000132109		2.50615/31913	302773		0.46800	10		In1/_mem	iory						
ENSG00000132400		2 1/00/20019	010024		0.51002	-		B-cell naive	0						
ENSG00000181099		1.9746408754	70262		-3 3279	2		NK-cell pai	ive						
ENSG00000180878		1.7731527333	799302		2.51023	-		monocyte	CD16 naive	a.					
Showing 1 to 10 of	292 entries								2	< < 1	2 3 4	5 > >>	1		
Regulatory featu	re consequences														
Show/hide colum	ns									Filter					
Regulatory feature	Active in cell line	98		Feature t	уре	Allele	Con	sequence ty	pe	.▲ Va	riant posi	tion	÷		
ENSR00000952768	IHECRE0000370	5, IHECRE0000	0989	promoter	flanking_region	А	reg	ulatory regior	n variant	26	01 (out of 45	98)			
ENSR00000952768	IHECRE00003705	5, IHECRE00000	0989	promoter.	_flanking_region	с 1 С 11	reg	ulatory regior	n variant	26	01 (out of 45	98)			
ENSR00000952768	Kegulatol	ry tea	tures	s the	varian	t falls	? 1 ¹ 499	ulatory regior	n variant	26	01 (out of 45	98)			

No overlap with Ensembl Motif features

This page illustrates the genes the variant falls within and the consequences on those genes, including pathogenicity predictors. It also shows data from GTEx on genes that have increased/decreased expression in individuals with this variant, in different tissues. Finally, regulatory features and motifs that the variant falls within are shown.

We can also see the variant in the protein structure by clicking on 3D Protein model.

3D Protein model (PDBe) @



This is a LiteMol viewer, where you can rotate and zoom in on the structure. The variant location is highlighted, so you can see where it lands within the structure.

Let's look at population genetics. Click on Population genetics in the left-hand menu.

Population genetics	0				
1000 Genomes Project Ph	nase 3 allele frequencies				
ALL	AMR	EAS	EUR	SAS	
• T: 97% • A: 3%	T: 90% A: 10% A: 1%	• T: 100%	+ T: 100%	+ T: 100%	
Sub-	populations Sub-populations	Sub-populations	Sub-populations E	Sub-population	v nie charte
Jump to: 1000 Genomes Proje Genome Variation Project (5)	ct Phase 3 (32) I gnomAD exomes	r <u>2.1.1 (9)</u> I <u>gnomAD geno</u>	mes r3.0 (10) I NCBI AL	FA.(10) I TOPMed (1) I NHLBI Exom	Sequencing Project (2) I Gambian
Hover over	population			_	
Show All v entries		Show/hide columns		Genotype	e frequencies
codes to ge	et their man	ves	Genoty	e: frequency (count)	Genotypes
ALL	T: 0.973 (4871)	A: 0.027 (137)	C TIT: 0.94	I5 (2367) AIT: 0.055 (137)	Show
AFR	T: 0.900 (1190)	A: 0.10 134 8 8 8	trequem	🖉 🌾 🖉 🖉 🖉 🖉 🖉 🖉 🖉 🖉	Show
inter 🖌 💻	T: 0.953 (183)	A: 0.047 (9)	TIT: 0.90	6 (87) AIT: 0.094 (9)	Show
Alrican	T: 0.984 (120)	A: 0.016 (2)	TIT: 0.96	67 (59) AIT: 0.033 (2)	Show
ESN	T: 0.879 (174)	A: 0.121 (24)	TIT: 0.75	8 (75) AIT: 0.242 (24)	Show

The population allele frequencies are shown by study, including 1000 Genomes and gnomAD. Where genotype frequencies are available, these are shown in the tables.

There are big differences in allele frequencies between populations. Let's have a look at the phenotypes associated with this variant to see if they are known to be specific to certain human populations. Click on *Phenotype Data* in the left-hand menu.

Variant rs334 (missense_variant) | Ensembl protein: ENSP00000333994 | PDBe model: 1DXT & (Hemoglobin subunit beta - chains B,D)

Phenotype Data 🚱

Significant association(s)

Show All 🗸 en	tries		Show/hide co	lumns				Filter	
Phenotype, disease and A trait	Source(s)	Mapped Terms	Ontology Accessions	Supporting evidence	External reference	Clinical significance	Reported gene(s)	Associated allele	Statistics
Anemia	ClinVar @ [New York Genome Cent]	Anemia, anemia (phenotype)	EFO:0004272년, HP:0001903년				LOC106099062, HBB, LOC107133510	Α	
<u>beta</u> Thalassemia	ClinVar® [MendelicsIMyriad Wom]	-	Orphanet:848			? tololok	LOC106099062, HBB, LOC107133510	Δ	
BETA- THALASSEMIA, DOMINANT INCLUSION BODY TYPE	ClinVart? [Division of Human Ge]		<u>Orphanet:231226</u> ස			A ****	LOC106099062, HBB, LOC107133510	Δ	
Blood cell traits multivariate analysis	NHGRI-EBI GWAS catalog P	11 terms ⊞	11 accessions ⊞	-	PMID:31080455@	-	HBB	-	p-value: 5.00e ⁻²⁰
Blood cell traits multivariate analysis	NHGRI-EBI GWAS catalog 관	11 terms ⊞	11 accessions ⊞		PMID:31080455@		HBB		p-value: 1.00e ⁻¹²

This variant is associated with various phenotypes, including sickle cell and malaria resistance. These phenotype associations come from sources including the GWAS catalog, ClinVar, Orphanet and OMIM. Where available, there are links to the original paper that made the association, the allele that is associated with the phenotype and p-values and other statistics.

Annotating genetic variants with the VEP, Demo

We have identified five variants on human chromosome nine, C-> A at 128203516, an A deletion at 128328461, C->A at 128322349, C->G at 128323079 and G->A at 128322917.

We will use the Ensembl VEP to determine:

- Have my variants already been annotated in Ensembl?
- What genes are affected by my variants?
- Do any of my variants affect gene regulation?

Go to the front page of Ensembl and click on the Variant Effect Predictor.

Variant Effect Predictor > Analyse your own variants and predict the functional consequences of known and unknown variants

Variant Effect Predictor @

This page contains information about the VEP, including links to download the script version of the tool. Click on *Launch VEP* to open the input form:

New job		Clear form	I Close
Species:	Homo_sapiens × Assembly: GRCh38 p13 Add/temoxe species If you are looking for VEP for Human GRCh37, please go to <u>GRCh37 website</u> \$0.		
Name for this job (optional):	~		
Input data:	Either paste data: Name your job Exemples: Ensemble debudt. VCP. Vorkent identifiers. HGVB notice Paste in your data Or upload file: Or provide file URL:		
Transcript database to use:	EnsembligENCODE transcripts Arrow unload a file		
	C Ensembl/GENCODE basic transmission C RefSeg transmission		
	· EnsomblyGENCOTE thoose with e transcript databa	se	

The data is in VCF format: chromosome coordinate id reference alternative

Put the following into the Paste data box:

9 128328460 var1 TA T 9 128322349 var2 C A 9 128323079 var3 C G 9 128322917 var4 G A 9 128203516 var5 C A

Additional configurations

The VEP will automatically detect that the data is in VCF.

There are further options that you can choose for your output. These are categorised as *Identifiers, Variants and frequency data, Additional annotations, Predictions, Filtering options* and *Advanced options*. Let's open all the menus and take a look.

Identifiers Additional identifiers for genes, transcripts and variants							
Identifiers							
Gene symbol:	۵						
Transcript version:	0						
0000	n						



BLOSUM62:						
Ancestral allele:	0					
Filtering options E Pre-filter results by frequency or	consequence type					
Filters						
Filter by frequency:	No filtering					
	Excluse common variants Advanced filtering					
Return results for variants in coding regions only:	• Filter the data					
Restrict results:	Show all results					
	NB: Restricting results may exclude biologically important data!					
Advanced options Additional enhancements						
	Run>					

Hover over the options to see definitions.

We're going to select some options:

- · HGVS, annotation of variants in terms of the transcripts and proteins they affect, commonly-used by the clinical community
- Phenotypes
- Protein domains

When you've selected everything you need, scroll right to the bottom and click Run.

Recent Jobs 🖻					
9 Refresh					
Show/hide columns (1 hidden)				Filter	
Analysis	Jobs		Submitted at		
Variant Effect Predictor	VEP analysis of pasted data in Homo_sapiens	Queued	17/09/2021, 14:39 (BST)		8/<1

The display will show you the status of your job. It will say *Queued*, then automatically switch to *Done* when the job is done, you do not need to refresh the page. You can edit or discard your job at this time. If you have submitted multiple jobs, they will all appear here.

Click View results once your job is done.

Variant Effect Predictor results @

In your results you will see a graphical summary of your data, as well as a table of your results.

Job details	E								
Category Variants prov Variants filte	cessed red out	Count 5 0	Consequences (all)	missense_variant: 32 upstream_gene_varia	% nt: 20%	Coding consequences	, ,		
Novel / exist Overlapped Overlapped Overlapped Results pre O Navigati Show: 1 5	ing variants genes transcripts regulatory feature view On (per variant) All variants	1 (20.0) / 4 (80.0) 4 24 25 2 b state V Filters Uploaded variant	istics	In-paneling site variant intron_variant: 11% regulatory_negion_vari downstream_gene_va non_coding_transcript non_coding_transcript non_coding_transcript NMD_transcript_variant: Filter	iant: 7% iant: 7% iriant: 5% Lexon_variant nt: 2% data Add	gene de Download All: VOE Y	nload y missens_variant: 5% wt your s into f New job	our a varia 3ioMi	lata unts/ art
Show/hide	o columns (23 hid	den)				BIOMART: Variants P	oenes &		
Uploaded variant	Location	Consequence	Symbol	Gene	Feature type	Feature	Biotype	Exon	HGVSc
var5	9:128203516- 128203516	missense_variant	DNM1	ENSG00000106976	Transcript	ENST0000034117 11	protein_coding	1/23	ENST0000341
var5	9:128203516- 128203516	missense_variant	DNM1	ENSG00000106976	Transcript	Detail	ed resul	lts"to	101000372
var5	9:128203516-	intron_variant	CIZ1	ENSG00000148337	Transcript	ENST00000372948.7	protein_coding	-	ENST00000372

The results table is enormous and detailed, so we're going to go through the it by section. The first column is *Uploaded variant*. If your input data contains IDs, like ours does, the ID is listed here. If your input data is

only loci, this column will contain the locus and alleles of the variant. You'll notice that the variants are not neccessarily in the order they were in in your input. You'll also see that there are multiple lines in the table for each variant, with each line representing one transcript or other feature the variant affects.

You can mouse over any column name to get a definition of what is shown.

The next few columns give the information about the feature the variant affects, including the consequence. Where the feature is a transcript, you will see the gene symbol and stable ID and the transcript stable ID and biotype. Where the feature is a regulatory feature, you will get the stable ID and type. For a transcription factor binding motif (labelled as a MotifFeature) you will see just the ID. Most of the IDs are links to take you to the gene, transcript or regulatory feature homepage.

Uploade d variant	Location 🔺	Consequence	Symbo I	Gene	Feature type	Feature	Biotype
var4	9:128322917- 128322917	missense_variant	COQ4	ENSG00000167113	Transcript	ENST00000372875.3	protein_coding
var4	9:128322917- 128322917	upstream_gene_variant	TRUB2	ENSG00000167112	Transcript	ENST0000372890.6	protein_coding
var4	9:128322917- 128322917	upstream_gene_variant	TRUB2	ENSG00000167112	Transcript	ENST00000460320.1	processed_transcript
var4	9:128322917- 128322917	missense_variant	COQ4	ENSG00000167113	Transcript	ENST0000608951.5	protein_coding
var4	9:128322917- 128322917	missense_variant	COQ4	ENSG00000167113	Transcript	ENST00000609948.1	protein_coding
var4	9:128322917- 128322917	regulatory_region_variant			RegulatoryFeature	ENSR00000241858	promoter
var4	9:128322917- 128322917	TF_binding_site_variant	-		MotifFeature	ENSM00000348409	
var4	9:128322917- 128322917	TF_binding_site_variant			MotifFeature	ENSM00150425749	

This is followed by details on the effects on transcripts, including the position of the variant in terms of the exon number, cDNA, CDS and protein, the amino acid and codon change, transcript flags, such as MANE, on the transcript, which can be used to choose a single transcript for variant reporting, and pathogenicity scores. The pathogenicity scores are shown as numbers with coloured highlights to indicate the prediction, and you can mouse-over the scores to get the prediction in words. Two options that we selected in the input form are the HGVS notation, which is shown to the left of the image below and can be used for reporting, and the Domains to the right. The Domains list the proteins domains found, and where there is available, provide a link to the 3D protein model which will launch a LiteMol viewer, highlighting the variant position.

Exon	HGVSc	cDNA position	CDS position	Protein position	Amino acids	Codons	Feature strand	MANE	Transcript support level	APPRIS	SIFT	PolyPhen	Domains
1/23	ENST00000341179.11:c.460-4	138	46	16	LM	CTG/ATG	1		1		0.02	0.01	Gene3D:3.40.50.300 PANTHER:PTHR11566 PANTHER:PTHR11566:SF32 SMART:SM00053 Superfamily:SSF52540
1/22	ENSTOCHT293	tati	ôn	16	LM	CTG/ATG	1	NM_004408.4	1	P4	0.02	0.015	Protein Structure www. 48 Protein domains @
-	ENST00000372948.7:c6+670G>T	· .	'n.		•	· .	-1	· .	2	A2	1)	in a state
1/23	ENST00000393594.7:c.46C>A	129	osi	tion	of va	ria	nt	N	IAN	E		Jami trui	PANTHER:PTHR11586 PANTHER:PTHR11586 PANTHER:PTHR11586:SF32 SMART:SM00053 Stearthroly:SSF52540
1/23	ENST00000475805.51c.46C>A	138	46	16	LM	CTG/ATC	5		2	A1	0.02	0.034	GenesD13.46.50.300 PANTHER:PTHR11566 PANTHER:PTHR11566:SF32 SMART:SM00053 Superfamily:SSF52540
1/22	ENST00000486160.3:c.46C>A	¹⁰² A	min	ö ac	id ar		òdo	n ch	ånge	Ŝ	P	atog	Gene3D:3.40.50.300 SMART:SM00352 Content:SST550 ANTHER:PTHR1566:SF32

Where the variant is known, the ID of the existing variant is listed, with a link out to the variant homepage. In this example, only rsIDs from dbSNP are shown, but sometimes you will see IDs from other sources such as COSMIC. The VEP also looks up the variant in the Ensembl database and pulls back the allele frequency (AF in the table), which will give you the 1000 Genomes Global Allele Frequency. In our query, we have not selected allele frequencies from the continental 1000 Genomes populations or from gnomAD, but these could also be shown here. We can also see ClinVar clinical significance and the phenotypes associated with known variants or with the genes affected by the variants, with the variant ID listed for variant associations and the gene ID listed for gene associations, along with the source of the association.

Existing variant	AF 🔶	Clinical significance	Associated phenotypes
rs9697215	0.0517	benign	ClinVar: phenotype not specified (<u>rs9697215</u> ,ClinVar)
rs9697215	0.0517	benign	ClinVar: phenotype not specified (<u>rs9697215</u> ,ClinVar)
<u>rs377735694</u>	0.0002	-	Neonatal encephalomyopathy- cardiomyopathy-respiratory distress syndrome (ENSG0000167113,Orphanet) COENZYME Q10 DEFICIENCY PRIMARY 7 (ENSG0000167113,MIM morbid & DDG2P) COENZYME Q10 DEFICIENCY PRIMARY 7 (rs377735694,ClinVar)

For variants that affect transcription factor binding motifs, there are columns that show the effect on motifs (you may need to click on *Show/hide columns* at the top left of the table to display these). Here you can see the position of the variant in the motif, if the change increases or decreases the binding affinity of the motif and the transcription factors that bind the motif.

Motif name	Motif position	High info position	Motif score change	TRANSCRIPTION FACTORS
ENSPFM0597	12	N	-0.005	TFAP2C::MAX
ENSPFM0177	1	N	-0.010	ETV6
ENSPFM0066	4	N	-0.055	E2F1::ELK1
ENSPFM0161	15	N	-0.019	ETV2::RFX5

Above the table is the *Filter* option, which allows you to filter by any column in the table. You can select a column from the drop-down, then a logic option from the next drop-down, then type in your filter to the following box. We'll try a filter of *Consequence*, followed by *is* then *missense_variant*, which will give us only variants that change the amino acid sequence of the protein. You'll notice that as you type *missense_variant*, the VEP will make suggestions for an autocomplete.

(G Filters					
R	 Uploaded variant 	is	✓ define	d		Add
	Location					
1	Allele					
ic	Consequence					
	Impact					
	Symbol	Amino acids	Codons	Existing	AF	clinic

You can export your VEP results in various formats, including VCF. When you export as VCF, you'll get all the VEP annotation listed under **CSQ** in the **INFO** column. After filtering your data, you'll see that you have the option to export only the filtered data. You can also drop all the genes you've found into the Gene BioMart, or all the known variants into the Variation BioMart to export further information about them.

Gene trees and homologues, Demo

Let's look at the homologues of human BRCA2. Search for the gene and go to the Gene tab.

Click on *Gene tree*, which will display the current gene in the context of a phylogenetic tree used to determine orthologues and paralogues.



Funnels indicate collapsed nodes. We can expand them by clicking on the node and selecting *Expand this sub-tree* from the pop-up menu.

Taxon: Laurasia (Laurasiatheria)	therian mammals
Gene Count	16
Branch Length	0.000505
Bootstrap	21
Туре	Speciation
Support	phyml-nt,phyml-aa
Image	expand this sub- tree
Image	expand all sub- trees
Image	collapse other nodes
node_id	49137553
Export sub- tree	Tree or Alignment
Export sub- tree	Sequences
View sub-tree	Wasabi viewer

We can also see the alignment of the sub-tree by clicking on Wasabi viewer, which will open a pop-up:



You can download the tree in a variety of formats. Click on the download icon in the bar at the top of the image to get a pop-up where you can choose your format.



File name:	BRCA2_gene_tree			
File format:	Choose Format	~		
	Preview Downlo	oad I Download Compressed		
Guide to file formats				
CLUSTALW	FASTA	Mega	MSF	
CLUSTAL W(1.81) multiple sequence home_sapiens/1-465588 CCTCAGAC pan_tregiotytes/1-465588 CCCCAGAC home_sapiens/1-465588 CCCAGTGCC pan_tregiodytes/1-465588 CCCAGTGCC	>homo_sapiens/1-464308 CCTCAGGACCGACGGCAAACCAACCAGAI CCCAGTACCTTCAGCTGCCTCTGGGCCC TGGGACAGGAGAAAACCACAGCTGGCT AGGGCCTGGTGGGGGCTGGTTGGGGGCT CCGAGGTGGATCCTGATATTGGCCACCTC CCGAGGTGGATCCTGATATTGGCCACCTC CCGGGCTGGGTGGCTGCTGGCTGTGTT AGGATGGGGGGGGGG	Mega ITille: ProjectedHultiAlign; Iformat datatyperdma identical=. Mhomo_sap CCTCA6GACC GACGGCAAAC Mhomo_sap CCCASTECT TCCACTCCT #homo_sap TCGGACAGAG AGAGAACAC	ProjectedMultiAlign MSF: 2 Type: Name: homo_sapiens/1-465588 Lee Name: pam_trogLodytes/1-465588 Lee // homo_sapiens/1-465588 CCTCAGGAC Go pum_trogLodytes/1-465588 CCTCAGGAC Go homo_sapiens/1-465588 CCTCAGGAC CI	

We can look at homologues in the *Orthologues* and *Paralogues* pages, which can be accessed from the lefthand menu. If there are no orthologues or paralogues, then the name will be greyed out. *Paralogues* is greyed out for *BRCA2* indicating that there are no paralogues available. Click on *Orthologues* to see the 175 orthologues available.

	Orthologues @								
	Download ortholog	lues					Ortholo	gue types	
	Summary of orthologues of this gene Hide \ominus								
	Click on 'Show details' to	display the orthologue	s for one or more gr	oups of species. Alternatively,	click on 'Configure this	s page' to choose	e a stat tom list of s	pecies.	
	Species set		Show details	With 1:1 orthologues	With 1:many ortholo	ogues Wit	h many:many orthologues	Without orthologues	
	Primates (26 species) Humans and other prima	ites		21	0		0	5	
	Rodents and related sp Rodents, lagomorphs and	cies (32 species) d tree shrews	2	22	1		0	9	
Choose a taxon	Laurasiatheria (45 spec Carnivores, ungulates an	cies) nd insectivores		36	1		0	8	
of interest	Placental Mammals (10 All placental mammals	18 species)		84	2		0	22	
	Sauropsida (69 species Birds and Reptiles)		23	1		0	45	
	Fish (86 species) Ray-finned fishes			50	1		0	35	
	All (280 species) All species, including inv	ertebrates		167	Infor	nation	on orth	oloques	
	Selected orthologues	Hide \ominus						č	
	Show All v entries		Sh	ow/hide columns			20	lter 🔝	
	Species 🔺	Туре	Orthologue		Target %id	Query G %id S	OC WGA core Cover	High age Confidence	
	Algerian mouse	1-to-1	Brca2 (MGP S	SPRETELJ G0029058)	58.86 %	57.31 %	99.35	Yes	
	(mus aprenus)	View Gene Tree	Compare Regio	ons (5:151,625,243-151,672,7	52:1)				
			View Sequence	Alignments					
	Alpine marmot	1-to-1	BRCA2 (ENS)	/MMG0000022103)	69.97 %	69.87 % 🚺	00] 100.0	0 Yes	
	(marmota)	View Gene Tree	Compare Regis 27,449,785:1)	ons (CZRN01000014.1:27,378	3,250-				

Choose to see only *Rodents and related species* orthologues by selecting the box. The table below will now only show details of rodent orthologues. Let's look at mouse.

Mouse	1-to-1	Brca2 (ENSMUSG00000041147)	58.58 %	57.05 %	100	0.00	Yes
(mus musculus)	<u>View Gene Tree</u>	Compare Regions (5:150,446,095- 150,493,794:1) View Sequence Alignments					

Links from the orthologue allow you to go to alignments of the orthologous proteins and cDNAs. Click on *View Sequence Alignments* then *View Protein Alignment* for the mouse orthologue.

Orthologue alignn Download homology Type: 1-to-1 orthologue	nent @ es				Infor ortho	mation on dogue pair
Species	Gene ID	Peptide ID	Peptide length	% identity (Protein)	% coverage	Genomic location
Human (Homo sapiens)	ENSG00000139618	ENSP00000369497	3418 aa	57 %	95 %	13:32315086-32400268
Mouse (Mus musculus)	ENSMUSG0000041147	ENSMUSP0000038576	3329 aa	58 %	98 %	<u>5:150446095-150493794</u>
ENSP00000369497/1-341 ENSMUSP00000038576/1-	8 MPIGSKERPTFFEIF 3329 MPVEYKRRPTFWEIF **: *.****:***	KTRCNKADLGPISLNWFEEI KARCSTADLGPISLNWFEEI *:*******************	LSSEAPPYNS LSSEAPPYNS	EPAEESEHKNNN EPPEESEYKPHG **.***:*:	YEPN ***:	
ENSP0000369497/1-341(ENSMUS 20000038576/1-3	8 LFKTPQRKPSYNQLA 3329 LFKTPQRNPPYHQFA ******:*.*:*:*	STPIIFKEQGLTLPLYQSP STPIMFKERSQTLPLDQSP ****:***:. **** ***	/KELDKFKLD FREL	DLGRNVPNSRHKS GKVVASSKHKI *: **:*:	ELRTV HSKK	
ENSP000003694977-341 ENSMUSP0000038576/1-3	8 KTKMDQADDVSCPLL 3329 KTKVDPVVDVASPPL ***:* . **:.* *	NSCLSESPVVLQCTHVTPQ KSCLSESPLTLRCTQAVLQ :******:.*:*:.*;	RDKSVVCGSI REKPVVSGSI *:*.**.**	FHTPKFVKGRQI FYTPKLKE-GQI *:***: : **	PKHI AI	ignment in Clustal W format
ENSP00000369497/1-341 ENSMUSP00000038576/1-3	8 SESLGAEVDPDMSWS 3329 SESLGVEVDPDMSWT *****.******	SSLATPPTLSSTVLIVRNEH SSLATPPTLSSTVLIARDEH *****************	EASETVFPHE EARSSVTPAE	DTTANVKSYFSNH SPATLKSCFSNH	IDESL INESP :**	

Whole genome alignments, Demo

Let's look at some of the comparative genomics views in the Location tab. Go to the region 2:176087000-176202000 in human, which contains the *HoxD* cluster which is involved in limb development and is highly conserved between species.

You can turn on conservation scores and constrained elements. Click on *Configure this page*, then *Comparative genomics* and turn on the tracks for *Constrained elements for 91 eutherian mammals EPO-Extended* and *Conservation score for 91 eutherian mammals EPO-Extended*. Save and close the menu.



You can now see the conservation scores in pale pink. These were used to determine the peaks indicated in the constrained elements track in dark pink. This track indicates regions of high conservation between species, considered to be "constrained" by evolution.

We can also look at individual species comparative genomics tracks in this view by clicking on *Configure this page*.

Select *BLASTz/LASTz* alignments from the left-hand menu to choose alignments between closely related species. Turn on the alignments for *Mouse*, *Chicken* and *Chimpanzee* in *Normal*. Save and close the menu.



The alignment is greatest between closely related species.

We can also look at the alignment between species or groups of species as text. Click on *Alignments (text)* in the left hand menu.

Select Select an alignment to open the alignment menu.

Select another alignment		
Alignments Selector	Selected species ¹	
Start typing the name of a species	Mouse reference (CL57BL6)	
All Alignments Pairwise Rodents & Lagomorphs Rats & Mice		
Reset All Cancel Apply		

Click through the links, *Pairwise*, *Rodents & Lagomorphs*, *Rats and Mice* to select *Mouse reference* (*CL57BL6*).

In this case there are two blocks aligned, Block 1 a large (115001 bp) alignment against mouse chr2 and one smaller block against mouse chr7. Click on *Block 1*.

You will see a list of the regions aligned, followed by the sequence alignment. Click on *Display full alignment*. Exons are shown in red.

To compare with both contigs visually, go to Region comparison.

To add species to this view, click on the blue *Select species or regions* button. Choose *Mouse Reference* again then close the menu.

🌣 🔝 < 🖽 🖃 🍳	
	119.60 kb. Forward strand -
Chr. 2 Genes (Comprehensive set from GENCODE 38)	176_1646 176_1246 176_1446 176_1646 176_1646 176_1646 176_2646 176_2646 176_2646
	Haxalis MR108 > Human region MR7704 >
Contigs Genes (Comprehensive set from GENCODE 38)	KC0121513 > KC012752 5
Regulatory Build	
Hsap-Mmus LastZ (o	Aligned regions are linked up
Che D	
Chr. 2	Pewerse strand 176.10Mb 176.12Mb 176.14Mb 176.16Mb 176.18Mb 176.20Mb
	»
	119.60 kb . Forward strand .
hsap-Mmus LastZ (o	74.50Mb 74.52Mb 74.54Mb 74.56Mb 74.56Mb 74.60Mb
Genes (Comprehensive set from GENCODE M	Hoxd13 > Hoxd12 > Gm28309 > Gm14396 > Hoxd8 > Hoxd3 > Hoxd3 > Hoxd12 > Hoxd10 > Hoxd10 > Hoxd4 > Hoxd1 > Hoxd1 > Hoxd10 > Hoxd4 > Hoxd1 > Hoxd
	Hoxd9 > Mr106 > Mr106 > Gm28230 >
Contigs Cones (Comprehensive	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
set from GENCODE M	< Gm28793 < Hagir
	< ENSMUSG00002076500
Regulatory Build	
Chr. 2	
	74.50Mb 74.52Mb 74.52Mb 74.58Mb 74.58Mb 74.60Mb

You can configure this view for both species. Click on *Configure this page* and look in the top left of the menu.

Species to configure:	✓ Human	1
	Mouse	5

The drop down allows you to configure each species separately.

We can view large scale syntenic regions from our chromosome of interest. Click on *Synteny* in the left hand menu.

Synteny @



Features that regulate gene expression, Demo

We're going to look for regulatory features in the region of a gene and investigate their activity in different cell types. We'll start by searching for the gene *KPNA2* and jumping to the **Location** tab. Scroll down to the **Region in detail** view and zoom out a little to see the gene as well as its flanking regions.

KPNA2-212 - ENST00000677086 > KPNA2-203 - ENST00000579754 NA2-206 - E 13302 -- **D** MANE Select Transcripts KPNA2-201 - ENST00000330459 : Regulatory Build . Contigs Age of Base Reg-features depicted as 68.05Mb 68.03Mb Reverse strand Variant Legend missense variant coloured blocks non coding transcript exon variant intergenic variant Structural Variant. CNV Gene Legend Protein Coding Non-Protein Coding merged Ensembl/Havana processed transcript Ensembl protein coding Regulation Legend CTCF enhancer promote promoter flank transcription factor binding Age of Base Legend Human-specific base eg-features (pale) = older) Appeared in mammals (paler = older) There are currently 871 tracks turned off Konsenbil Homo sapiens version 111.38 (GRCh38.p14) Chromosome 17: 68.029,772 - 68.053,229
 K Add/remove tracks | 👔 Custom tracks | < Share | 🖶 Resize image | 🔤 Export image | < Reset configurat

The Regulatory Build track is shown by default.

In this region we can see a number of regulatory features, including a red promoter with light red promoter flanks, cyan CTCF binding sites, yellow enhancers and lilac transcription factor (TF) binding sites (don't worry if you have zoomed out further or not as far and can see more/less). Refer to the legend at the bottom of the view to see what each of the colours mean.

You can also click on the individual regulatory features to learn more. Click on the red promoter to open a pop-up menu.



Click on the stable ID, ENSR0000097453, to jump to the Regulation tab.



Here, you can find a summary of the activity of the promoter in the different cell types. Scroll down to **Summary of Regulatory Aactivity** to find out in which cells the promoter is active (the feature displays an active epigenetic signature, which can include evidence of open chromatin), inactive (the region bears no epigenetic modifications from the ones included in the Regulatory Build), poised (the feature displays a epigenetic signature with the potential to be activated) or repressed (the feature is epigenetically suppressed). We can see that this promoter is active in one out of the 118 cell types currently in Ensembl.

Let's switch back to the **Location** tab to explore the different regulation tracks that are available. Click on **Configure this page** and in the pop-up window under the **Regulation** section, click on **Other regulatory regions** and enable the **Fantom 5**, **TarBase** and **Motif features** tracks. Close the pop-up window.

The **Fantom 5** track displays transcription start site (TSS) and enhancer predictions from the <u>FANTOM5</u> project.

The TarBase track displays experimentally verified miRNA targets from TarBase.

The **Motif features** track indicates the positions of transcription factor binding motifs (TFBMs) in black lines/blocks. You can click on individual features to find out more information about the TFBM, including a list of TFs binding at this site and, if available, in which cells the TFBM was experimentally verified in. You can also view the **Binding matrix**** by clicking on the matrix ID. This opens a pop-up window which displays the binding matrix used and a binding score representing how well a particular site matches the binding matrix.



We can explore more detailed data by adding further **Regulation** tracks. Click on the **Configure this page** button on the left-hand side.



In the pop-up window, go to **Regulation** and click on **Features by Cell/Tissue** to view the detailed activity of the regulatory feature by cell type.

n tracks					
Configure Region Image Configure Overview Image	Configure Chromosome Image Personal Data				
Active tracks					
this Favourite tracks	Select from available configurations: C	Current unsaved V Seve current configuration	0		
Track order					
Search results	Features by Cell/Tissue 1 Select tr	acks 2 Configure track display	3 View tracks		
JASPAR TFBS (0/2)					
Genome Reference Consortium Issues (0/17)	Cell/Tissue Find	Experiments		Selected tracks	Reset all
E Sequence and assembly (3/18)					
- Markers (0/1)			Select all Deselect all		
- GRC alignments (1/2)				Cell/Tissue	
- Simple features (0/4) Clones & miss, regions (0/7)	🛛 🗛 C D E F G H 👝 J K L I	м N O P O R S T U 👝 W 🗸 Y Z 🕨		Cell/Tissue 2 / 136 available	Ξ
E Genes and transcripts (2/79)	Coloor	t Coloat	aboa: Eia	aorta	
- Genes (2/7)	261661		pecific	astrocyte	
- Long reads (0/14)	A673				
 Prediction transcripts (0/1) LBG (0/1) 	adrenal cland			Experiments	
RNASeq models (0/56)		ue experi	mencs		
mRNA and protein alignments (0/7)	aorta			Histone 9 / 9 available	+
- mRNA alignments (0/3)	astrocyte			Open chromatin 1 / 1 available	+
Protein alignments (0/3)	astrocyte cerebellum			Transcription factors 2 / 2 available	+
E Variation (0/88)	 astrocyte spinal cord 				
- Sequence variants (0/20)					
 Phenotype, disease and curated variants (0/20) 					
- Array and other (0/12)	a state the for soll of				
View accailed	accivicy by cell c	ype			
 Somatic mutations (0/5) 					
- Somatic variants (0/2)					
- Somatic structural variants (0/3)					
Regulatory Build (1/1)					
Features by Cell/Tissue (18/415)	«				
DNA methylation (0/47) Other regulatory regions (2/4)					
E Comparative genomice (2/14)					
Multiple alignments (0/5)				Configure trac	ck display
- Conservation regions (2/5)					
 BLASTz/LASTz alignments (0/181) 					
Genome targeting (0/1)					
Oligo probac (0.96)					

We can add cells by clicking on them. Find them using the search or the alphabet ribbon. Let's add a cell type where the promoter is inactive, **aorta**, and one where it's active **astrocytes**. Once you've selected the cells, they will appear in the menu on the right, where you can easily view the list by clicking on the + icon and de-select them.

To choose the experiments to see data on, click on the **Experiments** tab at the top of the menu. You can navigate this the same as the **Cell/Tissue** tab, except that you have to choose between **Histone**, **Open Chromatin** and **Transcription factors**. Let's **Select all** in all categories.

When you've chosen your experiments and cells, you can click on the green **Configure track display** button in the bottom right-hand corner.



Now we can see the active feature in astrocytes compared to the inactive feature in aorta.



Bulk export of data with BioMart, Demo

Follow these instructions to guide you through BioMart to answer the following query:

You have three questions about a set of human genes: **ESPN, MYH9, USH1C, CISD2, THRB, WHRN** (these are HGNC gene symbols. More details on the HUGO Gene Nomenclature Committee can be found on http://www.genenames.org)

1. What are the NCBI Gene IDs for these genes?

2. Are there associated functions from the GO (gene ontology) project that might help describe their function?

3. What are their cDNA sequences?

Click on BioMart in the top header of a www.ensembl.org page to go to: www.ensembl.org/biomart/martview

You cannot choose any filters or attributes until you've chosen your dataset. Your dataset is the data type you're working with. In this case we're going to choose human genes, so pick *Ensembl Genes* then *Human genes* from the drop-downs.

CENSEMBI BLAST/BLAT VEP Tools BioMart Downloads Help & Docs Blog					
New Count Results		🚖 URL			
Dataset [None selected]	✓ - CHOOSE DATABASE - Ensembl Genes 104 Mouse strains 104 Ensembl Variation 104 Ensembl Regulation 104				
New Count Results		THE URL			
Dataset [None selected]	Ensembl Genes 104 - CHOOSE DATASET - Chicken genes (GRCg6a) Human genes (GRCh38,p13) Mouse genes (GRCm39)				

Now that you've chosen your dataset, the filters and attributes will appear in the column on the left. You can pick these in any order and the options you pick will appear.

Click on *Filters* on the left to see the available filters appear on the main page. You'll see that there are loads of categories of Filters to choose from. You can expand these by clicking on them. For our query, we're going to expand *GENE*.

Ner Count Results	Filterc 🔹 🖬	Perl ® Help
Dataset Human genes (G5 Ch38.p13)	Please restrict yo (If filter values are truncated in any	ar query using criteria below ists, hover over the list item to see the full text)
Filters	E REGION:	
[None selected]	e GENE:	
Attributes	Limit to genes external references)	With BioGRID Interaction data. The General Repository for Interaction Datasets
Gene stable ID	C Link to gones (when a references)	Only
Gene stable ID version		OExcluded
Transcript stable ID	Eunoral SENE	
Transcript stable ID version	Input e terra Krykin Aca ADD t [Nay 300 a week	Gene stable ID(s) [e.g. ENSG0000000003]

Our input data is a list of identifiers, so we're going to use the *Input external references ID list* filter. This allows us to input a list of identifiers from different databases. We need to choose what kind of identifier we're using, so that BioMart can look up the right column in a data table. You can pick these from a drop-down list, which lists the type of identifier with an example of how it looks. For our query, we have a list of gene names, so we need to pick *Gene Name*(s).



To check if the filters have worked, you can use the *Count* button at the top left, which will show you how many genes have passed the filter. If you get 0 or another number you don't expect, this can help you to see if your query was effective.



To choose the attributes, expand this in the menu. There are six categories for human gene attributes. These categories are mutually exclusive, you cannot pick attributes from multiple categories. This means that we need to do two separate queries to get our GO terms and NCBI IDs, and to get our cDNA sequences.



The Ensembl gene and transcript IDs, with and without version numbers are selected by default. The selected attributes are also listed on the left.

Filters	
Gene Name(s) [e.g. MT-TF]: [ID-list specified]	Features Variant (Germline) Structures Arionpilogues (Max sevel Collinologues) Germline
Gene stable ID Gene stable ID version Transcript stable ID Transcript stable ID version	GENE: Selected by default Gene stable ID Gene stable ID version Transcript stable ID
Dataset	Transcript stable ID version

We can choose the attributes we want by clicking on them. For our query, we're going to select:

GENE

- Gene Name
- EXTERNAL
 - NCBI gene ID
 - GO term accession
 - GO term name
 - GO term definition

We need to select the *Gene Name* in order to get back our original input, as this is not returned by default in BioMart. The order that you select the attributes in will define the order that the columns appear in in your output table.

You can get your results by clicking on Results at the top left.



The results table just gives you a preview of the first ten lines of your query. This allows the results to load quickly, so that if you need to make any changes to your query, you don't waste any time. To see the full table you can click on *View ## rows*. You can also export the data to an xls, tsv, csv or html file. For large queries, it is recommended that you export your data as *Compressed web file (notify by email)*, to ensure your download is not disrupted by connection issues.

Export all results	to	F	ile		TS	TSV V Dunique results only GO		
Email notification to								
View To view as HTML ~ Exports of ata						todata		
Gene stable ID	Transcript stable	Gene name	NCBI gene (formerly Entrezgene) ID	GO term accession	GO term name	GO term definition		
ENSG000010-34	walld	ato	<u>4627</u>	GO:0005524	ATP binding	Interacting selectively and non-covalently with ATP, adenosine 5'-triphosphate, a universally important coenzyme and enzyme regulator.		
ENSG0000100345	ENST00000216181	MYH9	4627	GO:0005515	protein binding	Interacting selectively and non-covalently with any protein or protein complex (a complex of two or more proteins that may include other nonprotein molecules).		
ENSG0000100345	ENST00000216181	MYH9	4627	<u>GO:0051015</u>	actin filament binding	Interacting selectively and non-covalently with an actin filament, also known as F-actin, a helical filamentous polymer of globular G-actin subunits.		
ENSG0000100345	ENST00000216181	MYH9	4627	GO:0003774	motor activity	Catalysis of the generation of force resulting either in movement along a microfilament or microtubule, or in torque resulting in membrane scission, coupled to the hydrolysis of a nucleoside triphosphate.		

You can see multiple rows per gene in your input list, because there are multiple transcripts per gene and multiple GO terms per transcript.

To get the cDNA sequences, go back to the *Attributes* then select the category *Sequences* and expand *SEQUENCES*.

When you select the sequence type, the part of the transcript model you've chosen will be highlighted in the grpahic.



Choose *cDNA* sequences, then expand *HEADER INFORMATION* to add *Gene Name* to the header. Then hit *Results* again.



For more details on BioMart, have a look at this publication:

Kinsella, R.J. et al Ensembl BioMarts: a hub for data retrieval across taxonomic space. http://europepmc.org/articles/PMC3170168

Custom data, Demo

Demo: Upload small files

We have some patients that present with microcephaly and developmental delay. They all have large scale deletions on chromosome five:

Patient	Chromosome	Start	End
P1	5	36821632	37091234
P2	5	36731476	36978306
P3	5	36908552	37108671

We can turn them into a BED file and view them in the genome browser:

chr5 36821632 37091234 P1 chr5 36731476 36978306 P2 chr5 36908552 37108671 P3

You can add data from a Region in Detail page by clicking on the *Custom tracks* button at the left. Alternatively, go to a species homepage and click on *Display your data in Ensembl*.

Display your data in Ensembl

A menu will appear:

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 part for more information.

Name for this data (optional):

Hume for th	io dua (optional).	
Species:	The species	Human (Homo sapidns) The data Assembly: GRCh38
Data:	is human	Paste in data or provide a file URL
		Paste in your data
		Or upload file (max 20MB) Choose File no file selected
Data format	t:	Help on supported formats, display types, etc
		Add data

The interface detects file types if you upload or attach a file. When you paste in your data, it can't do this so we have to tell it what our file type is. It will give you an option where you can select *BED*.

Click Add data.

You should get to a dialogue box telling you your upload has been successful.

Thank you. Your file uploaded successfully

File uploaded: BED demo (Bed file, Homo sapiens)

Total features found: 4

Go to Location view:

Nearest region with data: <u>5:36731477-36978306</u>

or

Close this window to return to current page

Click on the genomic coordinates link to go to the nearest region with data.

🌣 🔝 < 🖽 🖬 🇞	P2				Drag/Select: \leftrightarrow 👖
Chromosocratiands BED demo	36.750Mb 36.775Mb 3	36.800Mb 36.825Mb	246.83 kb 36.850Mb 36.875Mb p13.2	36.900Mb 36.9	Forward strand -
NAME AND ADDRESS	P2	Pl			
Genes (Comprehensiv set from GENCOLES8	ick on the	track	NIPE	The da	ta in the
n	ame to che	nae its	NIPE	L-202 > ein coding	
			pre	KRT18P31-201 > processed pseudogene	NIPBL-205 > processed trans
af	ppearance				NIPBL-204 : processed b
Genes (Comprehensive set from GENCODE 38)	AC026463.5 >	< AC026741.6	< AC026741.6 D0 < ENST00000 IncRNA	< ACC 605892	226741.6
			NIPBL-DT- IncRNA	201	
			< NIPBL IncRNA D-0 < NIPBL	-DT-203 L-DT-202	

To have a look at the file, click on *Custom tracks*.

Your data 🕖

Update	e selected	d: 💋 Connect 🛛 💋	Disconnect	Delete		
						Filter
Select	Туре	Source	Species	Asse mbly	Last updat ed	Actions
	Upload	BED demo View sample location Bed file	Homo sapiens Disc	GRCh38	17/11/21 at 14:10 ect ,	save, share
🛓 Ad	ld more data	Q Search for publ	ic track hu	lelete	e thi	s data

If you've got an Ensembl account, you can save this data to your account. Accounts are free to set up and allow you to save configurations and data, and share with groups.

Demo: Attach URLs of large files

Larger files, such as BAM files generated by NGS, need to be attached by URL. I've put a BAM file of human chromosome 20 RNASeq data online at: http://ftp.ebi.ac.uk/pub/databases/ensembl/training/emily_BAM/

Let's take a look at the folder.

Index of /pub/databases/ensembl/training/emily_BAM

Name	Last modified	Size Description
Parent Directory		-
GRCh37 Illumina reads test.bam	2019-01-16 10:01	4.3M
GRCh37 Illumina reads test.bam.bai	2019-01-17 14:10	176K
GRCh38.20.illumina.merged.1.bam	2019-01-17 14:09	2.8G
GRCh38.20.illumina.merged.1.bam.bai	2019-01-17 13:55	169K
GRCh38.21.illumina.merged.1.bam	2019-01-17 14:22	2.9G
GRCh38.21.illumina.merged.1.bam.bai	2019-01-17 14:11	121K
cat_e2.bam	2019-01-16 10:01	5.8M
2 cat e2.bam.bai	2019-01-16 09:17	131K

Apache Server at ftp.ebi.ac.uk Port 80

Here you can see a number of BAM files (.bam) with corresponding index files (.bam.bai). We're interested in the files *GRCh38.20.illumina.merged.1.bam* and *GRCh38.20.illumina.merged.1.bam.bai*. These files are the BAM file and the index file respectively. When attaching a BAM file to Ensembl, there must be an index file in the same folder.

To attach the file, click on *Custom tracks*, then click on *Add more data* to add a new track.

We get to the same dialogue box as before. This time we'll name our data Illumina reads.

Paste in the URL of the BAM file itself

(http://ftp.ebi.ac.uk/pub/databases/ensembl/training/emily_BAM/GRCh38.20.illumina.merged.1.bam).

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 <u>support page</u> for more information.

Name for this data (optional):	BAM demo
Species:	Human (<i>Homo sapiens</i>) Assembly: GRCh38
Data:	http://ftp.ebi.ac.uk/pub/databases/ensembl /training/emily_BAM/GRCh38.20.illumina.mer ged.l.bam
	Or upload file (max 20MB) Choose file No file chosen
Data format:	BAM Help on supported formats, display types, etc
	Add data

Since this is a file, the interface is able to detect the ".BAM" file extension, so automatically labels the format as BAM. Click on *Add data* and close the menu.

To see this data, jump to a region on chromosome 20. Let's go to the region of the CDH22 gene. Search for the gene and click on the location.



We can zoom in to see the sequence itself. Drag out boxes in the view to zoom in, until you see a view like this. Alternatively, jump to location 20:46241000-46241030.

Demo: Track hub registry

Our regulatory data incorporates data from sources such as ENCODE, Blueprint, and Roadmap Epigenomics. To see the data directly from these sources, you can add track hubs.

You can search for track hubs to add in different ways:

• Search for track hubs in the <u>Track Hub Registry</u> and choose to add them to your genome browser of choice.

• Search the track hub registry using the Track Hub Registry interface in Ensembl (there is a link from the homepage).

We will now add the track hub containing data from the Blueprint project.

You can add track hubs to view in Ensembl directly via the Track Hub Registry. Go to the <u>Track Hub Registry</u> <u>homepage</u> and search for *blueprint*.

The Track Hub Registry Submit data Docur	mentation - About	Help	Enter the search terms. Q	Sign up	Login
The J A global centra The goal of the Track Hub Registry i around the world to discov	Frack alised collect is to allow third pa ver and use track t	Hub Re ion of publicly access rities to advertise track hubs, a hubs containing different type	gistry ible track hubs and to make it easier for rese s of genomic research data.	archers	
Search by keywords: hg19, epigenomics, n	nouse			Q	
	Searc	ch box			
🛈 Submit Data 🔫	Q Access	s Data	III Stats		
I want maximum visibility for my rack hubs. External track hub providers can register and submit their track databases to the registry. Registration is web- based and done on this site; submission happens programatically via our REST with our combinition successfully validated, the travel of subme available search by other users workwide, allowing for automatic and rapid integration into a genome browser.	How do I f assembly of I Own to Control Track hubs can I information. Free ter box in the header of walk before a to Help on Advance	ind omics tracks for an my favourte organisty Subband be searched based on metadata ext search is provided from the search fail nack hub Registr web pages and other and subband to search and the search sed Search	A brief summary of the data co column for exact numbers.	Assemble	over the

There are two results for the Blueprint Hub, one for adding the track hub to GRCh37 and one for adding it to GRCh38, plus one RNA-seq alignment hub.



Alternatively, you can add track hubs by searching the Track Hub Registry through Ensembl. Click the *Custom tracks -> Track Hub Registry Search* in any region view within Ensembl.

Configure Region Image Configure Overview Image Configure Chromosome Image	Personal Data
Track Hub Registry Search Manage Configurations Succies:	Human (Homo sapiens)
Select Irack Hub Kegistry	GRCh38
search on the left	all V
Text search:	•
Cearch box	Hint: Leave "text search" empty to show all track hubs for this species
Search bux	Search

You can only find track hubs for the selected species and assembly denoted in the search box.

Search for blueprint.

Click Attach this hub in the search results page.

Configure Region Image Configure	e Overview Image Configure Chromosome Image	Personal Data
Custom tracks Track Hub Registry Search	Search Results	
Manage Configurations	Searched Human GRCh38 for "blueprint"	Can't see the track hub you're interested in?
	Found 1 track hub - <u>Search again</u>	We only search for hubs compatible with assemblies used on this website - please <u>search the registry directly</u> of for data on other assemblies. Alternatively, you can <u>manually attach any hub</u> for which you know the URL.
	Blueprint Hub Description: Blueprint Epigenomics Data Hub Data type: genomics Number of tracks: 5698	Attach this hub

Track Hubs often contain vast amounts of data, which can slow Ensembl down, so only add them if you need them, and trash them when you are finished with them.

Go to Configure this Page to see that a new category has been added to your menu.

Configure Region Image Configure Overview Image	e Configure Chromosome Image Personal Data				
Active tracks Favourite tracks	Select from available configurations:	Current unsaved V Se	ave current configuratio	n	
Search results	Blueprint Region Select tracks	2 Refine selection		Configure track display	w tracks
Genome Reference Consortium Issues (Q/17) Blueprint Hub (4/5695) Blueprint Region (3/2205) - Blueprint Signal (13/2490)	Sample description Find	Experiment Select all 1	Deselect all	Selected tracks	Reset all
Sequence and assembly (3/18) - Sequence (2/4) - Markors (0/1) - GRC alignments (1/2) - Simple features (0/4) - Clones & misc. regions (0/7)	Activated_B-Cell-Li Acute_F Acute_Lymphocyti Acute_F Acute_Lymphocyti Acute_F Acute_Wrebioi_Le Acute_F	romyelocy Germinal_Center [romyelocy Lymphoma_Follicular [romyelocy Mantie_Cell_Lymp fromyelocy Multiple_Myeloma	cytotoxic_0 effector_m effector_m	Sample description Sample description 1 / 96 available Experiment	٠
Genes and transcripts (2/77) Genes (2/5) Long reads (0/14) Prediction transcripts (0/1) LRG (0/1) RNASeq models (0/58)	Acute_Myeloid_Le Burkitt_i Acute_Myeloid_Le CD14-pt Acute_Myeloid_Le CD3-pet Acute_Premyelocy CD3-het Acute_Premyelocy CD3-het	Lymphoma Sporadic_Burkitt pative_C T-cell_Acute_Lymp gative_C T-cell_rolymphoc sitive_CD Type_1_diabetes gative_CC adult_enddhelial	endothelial endothelial erythroblas germinal_c hematopol	Experiment 7 / 9 available	۲
mRNA and protein alignments (2/7) - mRNA alignments (2/3) - EST alignments (0/1) - Protein alignments (0/3)	Acute_Promyelocy CD38-n Acute_Promyelocy CD4-poi Acute_Promyelocy CD4-poi Acute_Promyelocy CD8-poi	egative_na alternatively_activa [sitive_alp band_form_neutro [sitive_alp central_memory_C [sitive_alp. central_memory_C	immature_ inflammatc macrophag macrophag		
Heritation (304) - Sequence variants (1/20) - Phenotype, disease and curated variants (1/20) - Arrays and other (0/1) - Failed variants (0/1) - Structural variants (1/26)	Acute_Promyslocy CD8-por	sitive_alp class_switchedm _Lymphocy conventional_dend	macrophag	Filte	er tracks
Somatic mutations O(2) Somatic variants O(2) Somatic structural variants O(3)					

You can add tracks to the Region in Detail view using the matrix.

ł	🌣 🖪 < 🕀 🖬 🍫	5			
Γ		45,44 kb			
	Characteria handa	64.00Mb	64.01Mb	64.	02Mb
1	Chromosome bands			q23.3	
1	Blueprint Signal				
-	100.00-	C0010K DNase wiggler CD14-positive, CD16-negative das	sical monocyte signal from NCML	s	
2	C0010K.DNase.wiggler. mono T=0d				
	-000				
3	100.00- C0010K K4me3	C0010K H3K4me3 MACS2 wiggler CD14-positive, CD16-ne	gative classical monocyte signal	from NCMLS	
-	wiggler.mono_T=0d	· · · · · · · · · · · · · · · · · · ·			
	0.00				
2	Blueprint Region				
÷,	mono_T=0d				
2					
	C0010K.K4me3.peak.	• •			
2	mono_T=0d	Tracha Guar + 40			
-		Constrain Concession 0 Lutor Manuals E Ext	ended		
- 1	90 way GERP elements	. I II			
	Human cDNAs				
-	(Rerseg/ENA)	Klupprint track 4	ub		
2		Showing all 11 patures - click to show fewer	u D		
	Genes (Comprehensive				
	set from GENCODE 38)	PRR 29-202 >			
		PRR29-203 >			
		protein coding			
-		PRR29-205 >			
		protein coding			
-		PRR 29-209 >			
		protein coding			
-					
		protein coding			