

▼ Practical Bioinformatics. Homework 2

▼ UCSC

The UCSC Genome Browser is developed by the UCSC (University of California Santa Cruz) Genomics Institute.

The UCSC Genomes Database is a great resource for annotations, regulation and variation and all kinds of data for a growing number of taxa.

Let's choose any human protein and find any info about this protein gene in the UCSC Genome Browser:

I have chosen *Hemoglobin subunit alpha (HBA1)*

https://en.wikipedia.org/wiki/Hemoglobin_subunit_alpha.

At first, let's choose a reference genome.

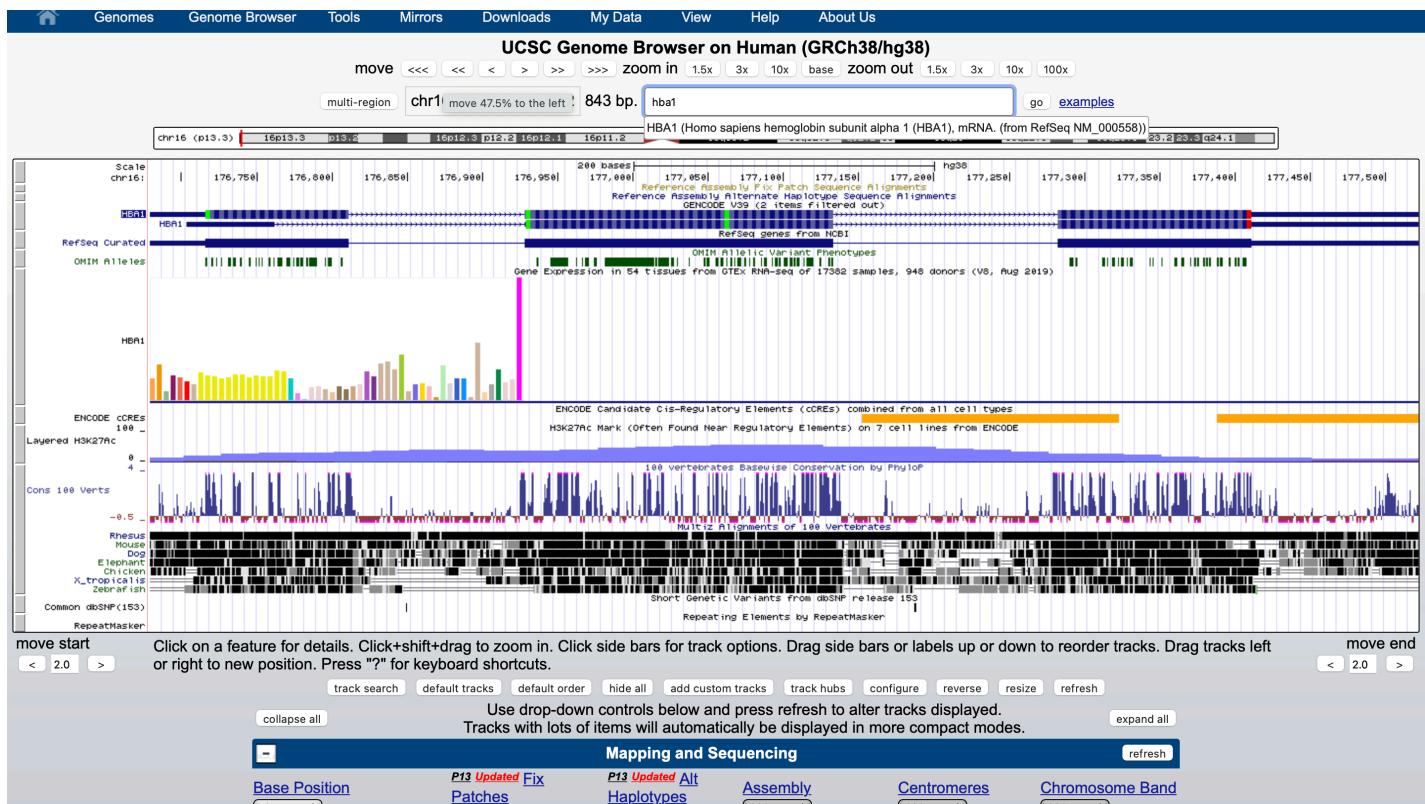
Some reference genome assemblies are available (at the UCSC website we may find - Human GRCh38/hg38, Human GRCh37/hg19). GRCh37 and hg19 are the same but named differently based on the institution, Genome Reference Consortium (GRC) and the University of California at Santa Cruz (UCSC), respectively.

The same is true of GRCh38 and hg38.

Both, GRCh37 and GRCh38 are human genome assemblies by the Genome Reference Consortium (GRC). GRCh38 (also called “build 38”) was released four years after the GRCh37 release in 2009, so it can be viewed as a version with updated annotations to the earlier assembly.

Let's select *hg38*.

Then we should see a view of the browser similar to the image below, opening at a position on the 16th chromosome of Human genome version GRCh38/hg38 showing the gene model for the HBA1 gene:



What it is needed to be highlighted:

- <> Gene name - *Hemoglobin subunit alpha (HBA1)*,
- <> Gene ID (from Gencode) - *ENSG00000206172.8* [you may see it on the image 2 below],
- <> Which strand is the gene encoded on / transcribed from? (+ or - strand)? - + [you may see it on the images 2,3 below: "Strand: +"],
- <> Chromosome - *chr16*,
- <> Which arm and position does this part of chromosome belong to? - *chr16:p13.3* (where "p" means the petit or short arm at the position 3 of chromosome 16. 13.3 - the number of a cytogenetic band). The red vertical bar shows where on the chromosome the current view is located. By the way, the designation q is for the long arm of the chromosome [you may see it on the image 1 above],
- <> How many alternative transcription products (alternative transcripts) may be encoded in the gene? (How many different transcripts variants are there for this gene?) - *According to the Gencode, this gene has 2 transcripts (so, 1 alternative transcript)*. Btw, according to the Ensembl Genome Browser it has 4 transcripts; but we need to use only the Gencode information (according to the homework instructions) [you may see it on the images 2-4 below],
- <> For each transcript it is needed to write its: Gencode ID, chromosome coordinates (including UTRs), number of exons (total), and length of a protein sequence (if there are more than three transcripts, then only the first three ones should be described):
 - 1) Gencode ID - *ENST00000320868.9*,
Chromosome coordinates (including UTRs) - *chr16:176,680-177,522 (hg38)*,
Number of exons - *Total Exon Count: 3*,
Length of a protein sequence - *Size: 142 aa*;
 - 2) Gencode ID - *ENST00000397797.1*,
Chromosome coordinates (including UTRs) - *chr16:176,704-177,522 (hg38)*,
Number of exons - *Total Exon Count: 3*,
Length of a protein sequence - *Size: 110 aa*.

Human Gene HBA1 (ENST00000320868.9) from GENCODE V39

Description: Homo sapiens hemoglobin subunit alpha 1 (HBA1), mRNA. (from RefSeq NM_000558)
RefSeq Summary (NM_000558): The human alpha globin gene cluster located on chromosome 16 spans about 30 kb and includes seven loci: 5'- zeta - pseudozeta - mu - pseudoalpha-1 - alpha-2 - alpha-1 - theta - 3'. The alpha-2 (HBA2) and alpha-1 (HBA1) coding sequences are identical. These genes differ slightly over the 5' untranslated regions and the introns, but they differ significantly over the 3' untranslated regions. Two alpha chains plus two beta chains constitute HbA, which in normal adult life comprises about 97% of the total hemoglobin; alpha chains combine with delta chains to constitute HbA-2, which with HbF (fetal hemoglobin) makes up the remaining 3% of adult hemoglobin. Alpha thalassemias result from deletions of each of the alpha genes as well as deletions of both HBA2 and HBA1; some nondeletion alpha thalassemias have also been reported. [provided by RefSeq, Jul 2008].

Gencode Transcript: ENST00000320868.9
Gencode Gene: ENSG00000206172.8
Transcript (Including UTRs)
Position: hg38 chr16:176,680-177,522 **Size:** 843 **Total Exon Count:** 3 **Strand:** +
Coding Region
Position: hg38 chr16:176,717-177,411 **Size:** 695 **Coding Exon Count:** 3

Page Index	Sequence and Links	UniProtKB Comments	MalaCards	CTD	RNA-Seq Expression
Microarray Expression	RNA Structure	Protein Structure	Other Species	GO Annotations	mRNA Descriptions
Pathways	Other Names	GeneReviews	Methods		

Data last updated at UCSC: 2022-01-17 08:30:34

- Sequence and Links to Tools and Databases

Genomic Sequence (chr16:176,680-177,522)	mRNA (may differ from genome)	Protein (142 aa)			
Gene Sorter	Genome Browser	Other Species FASTA	Gene interactions	Table Schema	BioGPS
CGAP	Ensembl	Entrez Gene	ExonPrimer	GeneCards	HGNC
HPRD	Lynx	MGI	neXtProt	OMIM	PubMed
Reactome	UniProtKB	Wikipedia			

- Comments and Description Text from UniProtKB

ID: HBA_HUMAN
DESCRIPTION: RecName: Full=Hemoglobin subunit alpha; AltName: Full=Alpha-globin; AltName: Full=Hemoglobin alpha chain;
FUNCTION: Involved in oxygen transport from the lung to the various peripheral tissues.
SUBUNIT: Heterotetramer of two alpha chains and two beta chains in adult hemoglobin A (HbA); two alpha chains and two delta chains in adult hemoglobin A2 (HbA2); two alpha chains and two epsilon chains in early embryonic hemoglobin Gower-2; two alpha chains and two gamma chains in fetal hemoglobin F (HbF).
INTERACTION: P68871:HBB; NbExp=19; IntAct=EBI-714680, EBI-715554;
TISSUE SPECIFICITY: Red blood cells.
PTM: The initiator Met is not cleaved in variant Thionville and is acetylated.
DISEASE: Defects in HBA1 may be a cause of Heinz body anemias (HEIRAN) IMIM-1407001 This is a form of non-spherocytic hemolytic anemia of Dacie type 1. After splenectomy, which has little

Human Gene HBA1 (ENST00000397797.1) from GENCODE V39

Description: Involved in oxygen transport from the lung to the various peripheral tissues (By similarity). (from UniProt G3V1N2)
RefSeq Summary (NM_000558): The human alpha globin gene cluster located on chromosome 16 spans about 30 kb and includes seven loci: 5'- zeta - pseudozeta - mu - pseudoalpha-1 - alpha-2 - alpha-1 - theta - 3'. The alpha-2 (HBA2) and alpha-1 (HBA1) coding sequences are identical. These genes differ slightly over the 5' untranslated regions and the introns, but they differ significantly over the 3' untranslated regions. Two alpha chains plus two beta chains constitute HbA, which in normal adult life comprises about 97% of the total hemoglobin; alpha chains combine with delta chains to constitute HbA-2, which with HbF (fetal hemoglobin) makes up the remaining 3% of adult hemoglobin. Alpha thalassemias result from deletions of each of the alpha genes as well as deletions of both HBA2 and HBA1; some nondeletion alpha thalassemias have also been reported. [provided by RefSeq, Jul 2008].

Gencode Transcript: ENST00000397797.1
Gencode Gene: ENSG00000206172.8
Transcript (Including UTRs)
Position: hg38 chr16:176,704-177,522 **Size:** 819 **Total Exon Count:** 3 **Strand:** +
Coding Region
Position: hg38 chr16:176,930-177,411 **Size:** 482 **Coding Exon Count:** 2

Page Index	Sequence and Links	UniProtKB Comments	MalaCards	CTD	RNA-Seq Expression
Microarray Expression	RNA Structure	Protein Structure	Other Species	GO Annotations	mRNA Descriptions
Pathways	Other Names	GeneReviews	Methods		

Data last updated at UCSC: 2022-01-17 08:30:34

- Sequence and Links to Tools and Databases

Genomic Sequence (chr16:176,704-177,522)	mRNA (may differ from genome)	Protein (110 aa)			
Gene Sorter	Genome Browser	Other Species FASTA	Gene interactions	Table Schema	BioGPS
CGAP	Ensembl	ExonPrimer	GeneCards	HGNC	Lynx
MGI	PubMed	UniProtKB	Wikipedia		

Additional information:

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

Human (GRCh38.p13) ▾

Location: 16:176,680-177,522 Gene: HBA1 Transcript: HBA1-201

Gene-based displays

- Summary
- Splice variants
- Transcript comparison
- Gene alleles
- Sequence
- Secondary Structure
- Comparative Genomics
- Genomic alignments
- Gene tree
- Gene gain/loss tree
- Orthologues
- Paralogues
- Ensembl protein families
- Ontologies
- GO: Biological process
- GO: Cellular component
- GO: Molecular function
- Phenotypes
- Genetic Variation
- Variant table
- Variant image
- Structural variants
- Gene expression

Gene: HBA1 ENSG00000206172

Description: hemoglobin subunit alpha 1 [Source:HGNC Symbol;Acc:HGNC:4823]

Gene Synonyms: HBA-T3

Location: Chromosome 16: 176,680-177,522 forward strand. GRCh38:CM000678.2

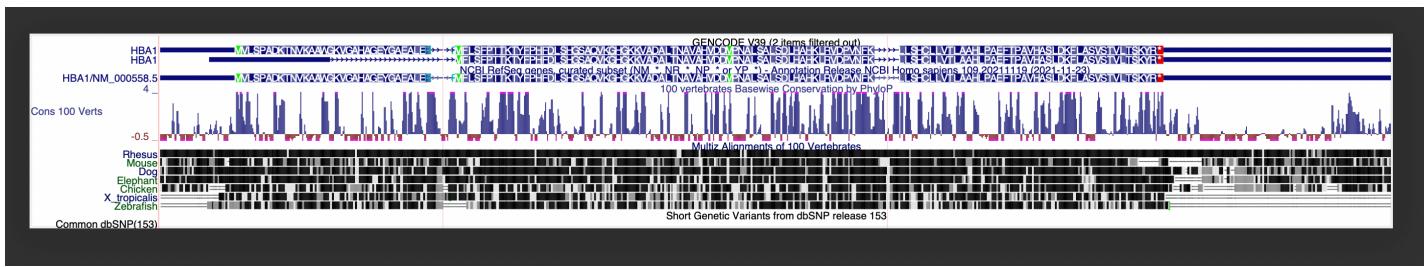
About this gene: This gene has 4 transcripts (splice variants), 230 orthologues, 12 paralogues and is associated with 9 phenotypes.

Transcripts: Hide transcript table

Show/hide columns (1 hidden) Filter

Transcript ID	Name	bp	Protein	Biotype	CCDS	UniProt Match	RefSeq Match	Flags
ENST00000320868.9	HBA1-201	577	142aa	Protein coding	CCDS10399	D1MGQ2 P69905	NM_000558.5	MANE Select v0.95 Ensembl Canonical GENCODE basic APPRIS P1 TSL:1
ENST00000397797.1	HBA1-202	504	110aa	Protein coding	-	G3V1N2	-	GENCODE basic TSL:2
ENST00000472694.1	HBA1-203	674	No protein	Retained intron	-	-	-	TSL:1
ENST00000487791.1	HBA1-204	410	No protein	Retained intron	-	-	-	TSL:1

Then, save a picture of the gene's locality from the UCSC Genome Browser with the tracks: GENCODE and RefSeq transcripts, conservation of the sequence among vertebrates, frequent polymorphisms (Common SNPs) from the latest version (dbSNP(153)) and hide all other tracks.



Mapping and Sequencing

Base Position	P13 Updated Fix	P13 Updated Alt	Assembly	Centromeres	Chromosome Band
hide ↕	hide ↕	hide ↕	hide ↕	hide ↕	hide ↕
Clone Ends	Exome Probesets	18 FISH Clones	Gap	GC Percent	GRC Contigs
hide ↕	hide ↕	hide ↕	hide ↕	hide ↕	hide ↕
GRC Incident	Hg19 Diff	INSDC	LiftOver & ReMap	LRG Regions	Mappability
hide ↕	hide ↕	hide ↕	hide ↕	hide ↕	hide ↕
RefSeq Acc	Restr Enzymes	Scaffolds	Short Match	STS Markers	
hide ↕	hide ↕	hide ↕	hide ↕	hide ↕	

Genes and Gene Predictions

GENCODE V39	NCBI RefSeq	New All GENCODE	CCDS	CRISPR Targets	19 IKMC Genes Mapped
full ↕	full ↕	hide ↕	hide ↕	hide ↕	hide ↕
LRG Transcripts	Updated MANE select v0.95	MGC Genes	Non-coding RNA	Old UCSC Genes	ORFeome Clones
hide ↕	hide ↕	hide ↕	hide ↕	hide ↕	hide ↕
Other RefSeq	Pfam in GENCODE	New Prediction Archive	RetroGenes V9	TransMap V5	UCSC Alt Events
hide ↕	hide ↕	hide ↕	hide ↕	hide ↕	hide ↕
UniProt					
hide ↕					

Phenotype and Literature

OMIM Alleles	CADD	Cancer Gene Expr	ClinGen	Deprecated ClinGen CNVs	ClinVar Variants
hide ↕	hide ↕	hide ↕	hide ↕	hide ↕	hide ↕
19 Coriell CNVs	COSMIC Regions	Development Delay	Gene Interactions	GeneReviews	GWAS Catalog
hide ↕	hide ↕	hide ↕	hide ↕	hide ↕	hide ↕
HGMD Variants	LOVD Variants	OMIM Cyto Loci	OMIM Genes	New Orphanet	19 New REVEL Scores
hide ↕	hide ↕	hide ↕	hide ↕	hide ↕	hide ↕
SNPedia	TCGA Pan-Cancer	UniProt Variants	Variants in Papers		
hide ↕	hide ↕	hide ↕	hide ↕		

COVID-19

[COVID GWAS v4](#) [COVID GWAS v3](#) [Rare Harmful Vars](#) [refresh](#)

[Colon Wang](#) [Ileum Wang](#) [Rectum Wang](#) [Blood \(PBMC\)](#) [Cortex Velmeshev](#) [Fetal Gene Atlas](#)
[hide](#) [hide](#) [hide](#) [hide](#) [hide](#) [hide](#)

[Heart Cell Atlas](#) [Kidney Stewart](#) [Liver MacParland](#) [Lung Travaglini](#) [Muscle De Micheli](#) [Pancreas Baron](#)
[hide](#) [hide](#) [hide](#) [hide](#) [hide](#) [hide](#)

[Placenta Vento-Tormo](#) [Skin Sole-Boldo](#)
[hide](#) [hide](#)

mRNA and EST

[Human ESTs](#) [Human mRNAs](#) [Other ESTs](#) [Other mRNAs](#) [SIB Alt-Splicing](#) [Spliced ESTs](#) [refresh](#)

[hide](#) [hide](#) [hide](#) [hide](#) [hide](#) [hide](#)

Expression

[GTEx Gene V8](#) [GTEx RNA-Seq Coverage](#) [Affy Archive](#) [EPDnew Promoters](#) [GNF Atlas 2](#) [GTEx Gene](#)
[hide](#) [hide](#) [hide](#) [hide](#) [hide](#) [hide](#)

[GTEx Transcript](#) [GWIPS-viz Riboseq](#) [miRNA Tissue Atlas](#)
[hide](#) [hide](#) [hide](#)

Regulation

[ENCODE cCREs](#) [ENCODE Regulation](#) [CpG Islands](#) [GeneHancer](#) [GTEx cis-eQTLs](#) [Hi-C and Micro-C](#)
[hide](#) [hide](#) [hide](#) [hide](#) [hide](#) [hide](#)
GTEX High-Confidence cis-eQTLs from CAVIAR

[New JASPAR Transcription Factors](#) [ORegAnno](#) [RefSeq Func Elms](#)
[hide](#) [hide](#) [hide](#)

Comparative Genomics

[Conservation](#) [Cactus 241-way](#) [Cons 30 Primates](#) [Primate Chain/Net](#) [Placental Chain/Net](#) [Vertebrate Chain/Net](#)
[full](#) [hide](#) [hide](#) [hide](#) [hide](#) [hide](#)

Variation

[dbSNP 153](#) [New 1000G Archive](#) [dbSNP Archive](#) [dbVar Common Struct Var](#) [DGV Struct Var](#) [Genome In a Bottle](#)
[full](#) [hide](#) [hide](#) [hide](#) [hide](#) [hide](#)

[gnomAD Variants](#) [Platinum Genomes](#)
[hide](#) [hide](#)

Repeats

[RepeatMasker](#) [Interrupted Rpts](#) [Microsatellite](#) [RepeatMasker Viz. Segmental Dups](#) [Self Chain](#)
[hide](#) [hide](#) [hide](#) [hide](#) [hide](#)

[Simple Repeats](#) [WM + SDust](#)
[hide](#) [hide](#)

▼ Ensembl

Ensembl is a genome browser for vertebrate genomes that supports research in comparative genomics, evolution, sequence variation and transcriptional regulation. Ensembl annotate genes, computes multiple alignments, predicts regulatory function and collects disease data. Ensembl tools include BLAST, BLAT, BioMart and the Variant Effect Predictor (VEP) for all supported species.

<https://www.ensembl.org/index.html>

Align the chosen human protein gene and the homologous chimpanzee protein gene: Find the specified gene in the Ensembl search. In the menu on the left side, select Comparative Genomics → Genomic alignments. Click Select another alignment, enter the name of the organism - Chimpanzee → Apply. Save the resulting alignment (Download).

The screenshot shows the Ensembl homepage. At the top, there is a dark blue header with the Ensembl logo and a navigation bar with links: BLAST/BLAT | VEP | Tools | BioMart | Downloads | Help & Docs | Blog. On the right edge of the header, there is vertical text: E n t i a E F. Below the header, there is a white navigation bar with four main sections: Tools, BioMart >, BLAST/BLAT >, and Variant Effect Predictor >. Each section has a link to its documentation. The main content area is titled "Search". It features a search input field with "Human" and "for" dropdown menus, and a dropdown menu showing suggestions for "hba1": "Search for 'hba1'" and "Search for 'hba1c'". Below the search input is a "Direct Links" section with a yellow-highlighted link to "HBA1" (ENSG00000206172 Gene ENSG00000206172) and "HbA1c measurement". To the right of the search area, there are three cards: "Human" (GRCh38.p13), "Mouse" (GRCm39), and "Zebrafish" (GRCz11). On the left, there is a "All genomes" section with a dropdown menu "Select a species --" and a "Pig breeds" section with a thumbnail of a pig and text "Pig reference genome and 12 additional breeds". At the bottom left, there is a link "View full list of all species".

Human (GRCh38.p13) ▾

Location: 16:176,680-177,522 Gene: HBA1

Gene-based displays

- Summary
- Splice variants
- Transcript comparison
- Gene alleles

Sequence

- Secondary Structure
- Comparative Genomics
- Genomic alignments**
- Gene tree
- Gene tree/loss tree
- Orthologues
- Paralogues

Ensembl protein families

Ontologies

- GO: Biological process
- GO: Cellular component
- GO: Molecular function

Phenotypes

Genetic Variation

- Variant table
- Variant image
- Structural variants
- Gene expression
- Pathway
- Regulation
- External references
- Supporting evidence

ID History

- Gene history

Gene: HBA1 ENSG00000206172

Description hemoglobin subunit alpha 1 [Source:HGNC Symbol;Acc:[HGNC:4823](#)]

Gene Synonyms HBA-T3

Location [Chromosome 16: 176,680-177,522](#) forward strand. GRCh38:CM000678.2

About this gene This gene has 4 transcripts ([splice variants](#)), [230 orthologues](#), [12 paralogues](#) and is associated with [9 phenotypes](#).

Transcripts [Hide transcript table](#)

Show/hide columns (1 hidden) Filter

Transcript ID	Name	bp	Protein	Biotype	CCDS	UniProt Match	RefSeq Match	Flags				
ENST00000320868.9	HBA1-201	577	142aa	Protein coding	CCDS10399	D1MQQ2	P69905	NM_000558.5	MANE Select v0.95	Ensembl Canonical	GENCODE basic	APPRIS P1
ENST00000397797.1	HBA1-202	504	110aa	Protein coding	-	G3V1N2	-	-			GENCODE basic	TSL:2
ENST00000472694.1	HBA1-203	674	No protein	Retained intron	-	-	-	-				TSL:1
ENST00000487791.1	HBA1-204	410	No protein	Retained intron	-	-	-	-				TSL:1

Genomic alignments [?](#)

Alignment: Chimpanzee [Select another alignment](#)

[Configure this page](#)

[Download alignment](#)

[View an image of this alignment](#)

A total of 2 alignment blocks have been found. Please select an alignment to view by selecting a Block from the Alignment column.

Show/hide columns Filter

Alignment (click to view)	Length (bp)	Location on Human	Location on Chimpzee
---------------------------	-------------	-------------------	----------------------

Then, we need estimate the percentage of differences per 100 nucleotides between these 2 alignments.

It can be done (to count the number of differences) by the command *infoalign* of EMBOSS package or with biopython.

```
!pip install biopython
```

```
Collecting biopython
  Downloading biopython-1.79-cp37-cp37m-manylinux_2_5_x86_64.manylinux1_x86
  |██████████| 2.3 MB 5.1 MB/s
Requirement already satisfied: numpy in /usr/local/lib/python3.7/dist-pac
Installing collected packages: biopython
Successfully installed biopython-1.79
```

```
from Bio import AlignIO
```

```
alignment = AlignIO.read(open('Human_Chimpanzee_HBA1.fa'), 'fasta')
seq1 = str(alignment[0].seq)
seq2 = str(alignment[1].seq)
matches = sum(nuc1 == nuc2 for nuc1, nuc2 in zip(seq1, seq2))
identity = 100.0 * matches / len(seq1)
identity # 87.4506
```

```
87.4505928853755
```

Compare the obtained value with the genome-wide assessment and specify the source:

From the article "Initial sequence of the chimpanzee genome and comparison with the human genome" (The Chimpanzee Sequencing and Analysis Consortium, Published: 1 September 2005, <https://www.nature.com/articles/nature04072>).

Section "Genome evolution. Nucleotide divergence" /

".. **Genome-wide rates. We calculate the genome-wide nucleotide divergence between human and chimpanzee to be 1.23%, confirming recent results from more limited studies.**.. The differences between one copy of the human genome and one copy of the chimpanzee genome include both the sites of fixed divergence between the species and some polymorphic sites within each species. By correcting for the estimated coalescence times in the human and chimpanzee populations.. we estimate that polymorphism accounts for 14–22% of the observed divergence rate and thus that the fixed divergence is ~1.06% or less.

Nucleotide divergence rates are not constant across the genome, as has been seen in comparisons of the human and murid genomes.. The average divergence in 1-Mb segments fluctuates with a standard deviation of 0.25% (coefficient of variation = 0.20), which is much greater than the 0.02% expected assuming a uniform divergence rate.."

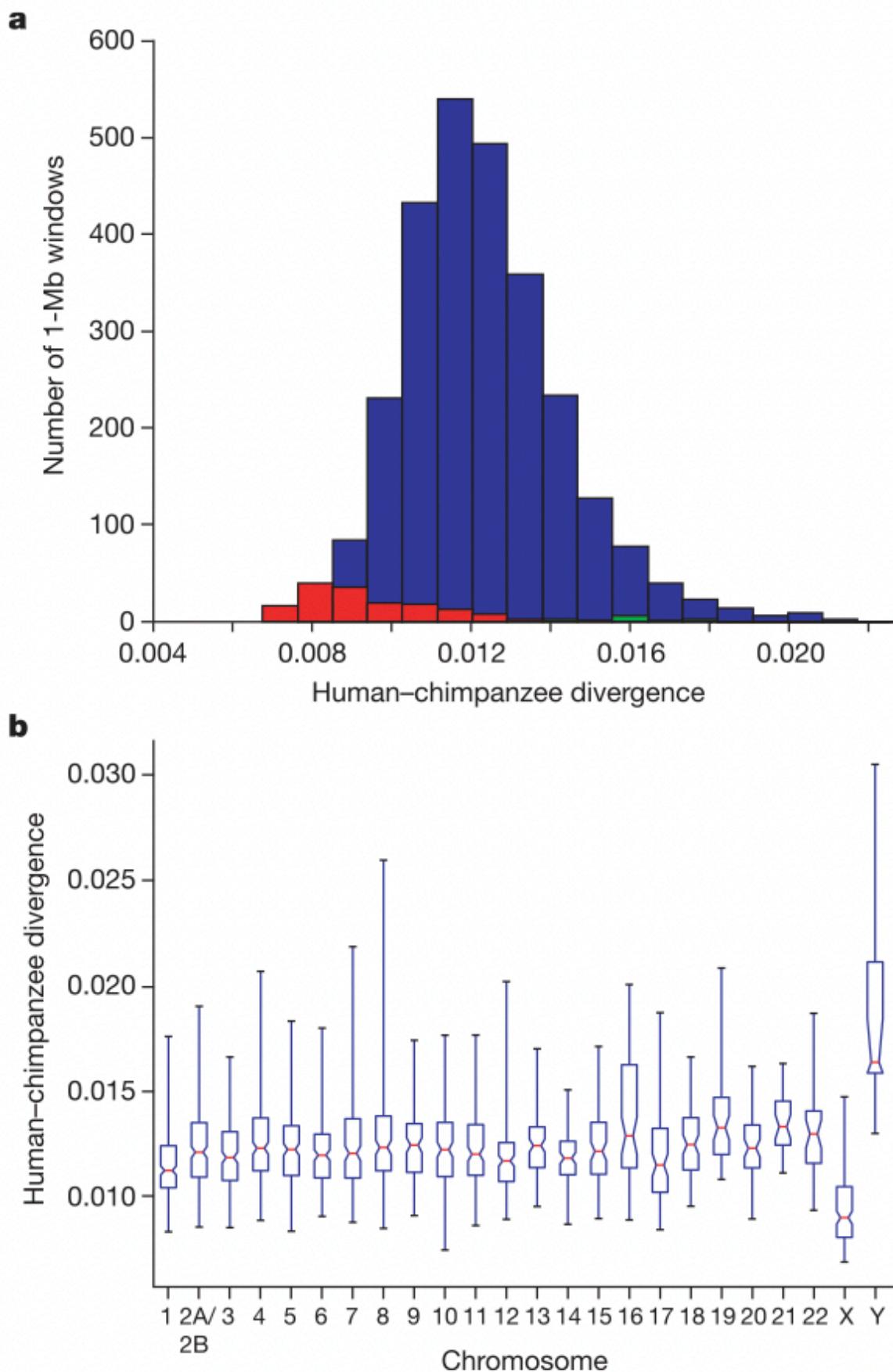


Figure 1 | Human-chimpanzee divergence in 1-Mb segments across the genome. a, Distribution of divergence of the autosomes (blue), the X chromosome (red) and the Y chromosome (green). b, Distribution of variation by chromosome, shown as a box plot. The edges of the box

variation by chromosome, shown as a box plot. The edges of the box correspond to quartiles; the notches to the standard error of the median; and the vertical bars to the range. The X and Y chromosomes are clear outliers, but there is also high local variation within each of the autosomes.

Divergence = 1.23

Identity_genome_wide = 100 - Divergence

Identity_genome_wide # 98.77 genome-wide assessment of identity of human and chi

98.77

The Percent Difference comparison calculates the percentage difference between two number values in order to determine how close they are.

The Percent Difference (PD) between two numbers is calculated as follows:

PD = $|(\text{n2}-\text{n1})/((\text{n1}+\text{n2})/2)| \times 100\%$, where ni - a number.

PD = ((Identity_genome_wide - identity)/((Identity_genome_wide + identity)/2))*1
PD # 12.157%

12.156987515974606

▼ Genome Data Viewer

the link to the Genome Data Viewer is <https://www.ncbi.nlm.nih.gov/genome/gdv/>.

Observe the same genome fragment in the Genome Data Viewer.

U.S. National Library of Medicine NCBI National Center for Biotechnology Information Log In

Genome Data Viewer

Switch view Search organisms Homo sapiens (human)

To view more organisms in the tree, click on nodes that have '+' signs. Press and hold the '+' to expand and reveal all the subgroups. Or, search for an organism using the search box above.

New! Click on Switch view at the top to see another way of navigating genomes.

Homo sapiens (human)

Search in genome... HBA1

Name	Location
HBA1	Chr16: 226,679 - 227,521
EGFR	Chr7: 55,086,710 - 55,279,321
LOC106804613	Chr16: 224,617 - 227,540
PTEN	Chr10: 89,623,382 - 89,731,687
ESR2	Chr14: 64,693,425 - 64,805,331
HBA2	Chr16: 222,875 - 223,709
APOA1	Chr11: 116,706,467 - 116,708,666
HBB	Chr11: 5,246,694 - 5,248,301

Examples: TP53, chr17:7667000-7689000, DNA repair

Assembly GRCh37.p13

Tip: You can filter RefSeq and Ensembl-annotated gene tracks to show only coding transcripts, hide gene models, and more. Click the gear icon to change track display settings.

Genome Data Viewer

Homo sapiens (human) Assembly: GRCh37.p13 (GCF_000001405.25) • Chr 16 (NC_000016.9)

Search assembly HBA1 Examples ▾

Pick Assembly Ideogram View User Data and Track Hubs BLAST

Select BLAST RID Tools Add Tracks by Accession Assembly Region Details History

NC_000016.9: 226,595 - 227,605

Gene Transcript Exons: click an exon to zoom in, mouse over to see details

Region HBA1 NM_000558.5

Tools Tracks Download ?

Genes, NCBI Homo sapiens Annotation Release 105.20201022

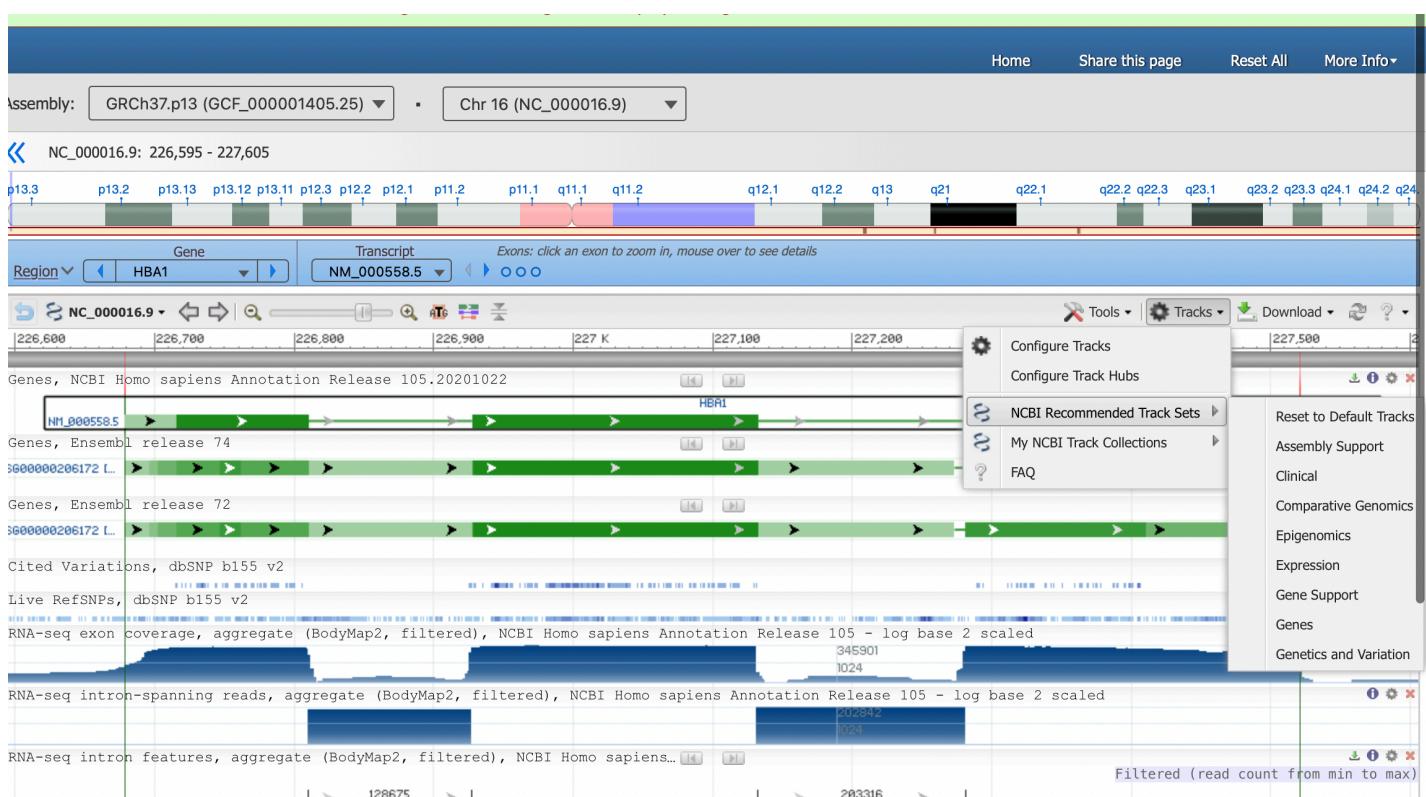
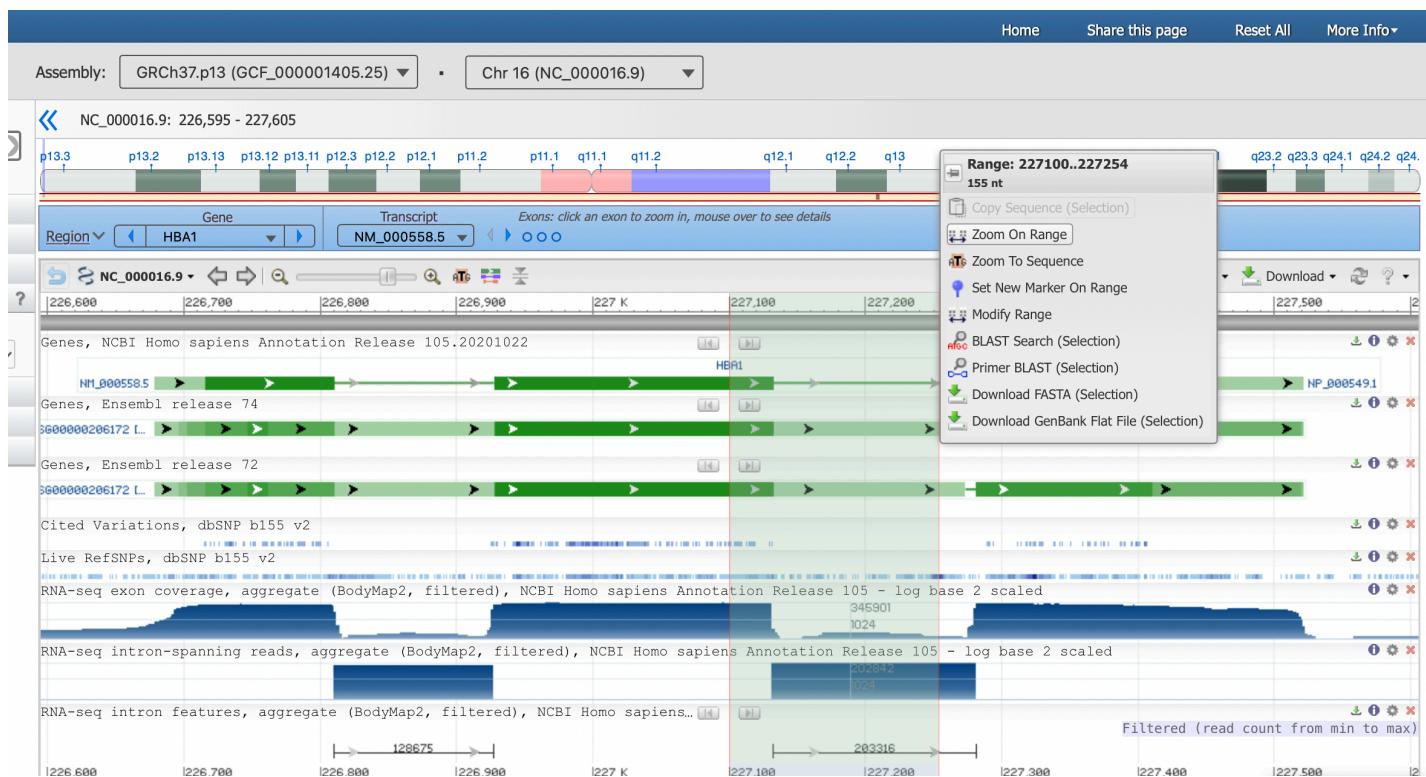
NM_000558.5 Genes, Ensembl release 74 Genes, Ensembl release 72 Cited Variations, dbSNP b155 v2 Live RefSNPs, dbSNP b155 v2 RNA-seq exon coverage, aggregate (BodyMap2, filtered), NCBI Homo sapiens Annotation Release 105 - log base 2 scaled RNA-seq intron-spanning reads, aggregate (BodyMap2, filtered), NCBI Homo sapiens Annotation Release 105 - log base 2 scaled RNA-seq intron features, aggregate (BodyMap2, filtered), NCBI Homo sapiens...

Filtered (read count from min to max)

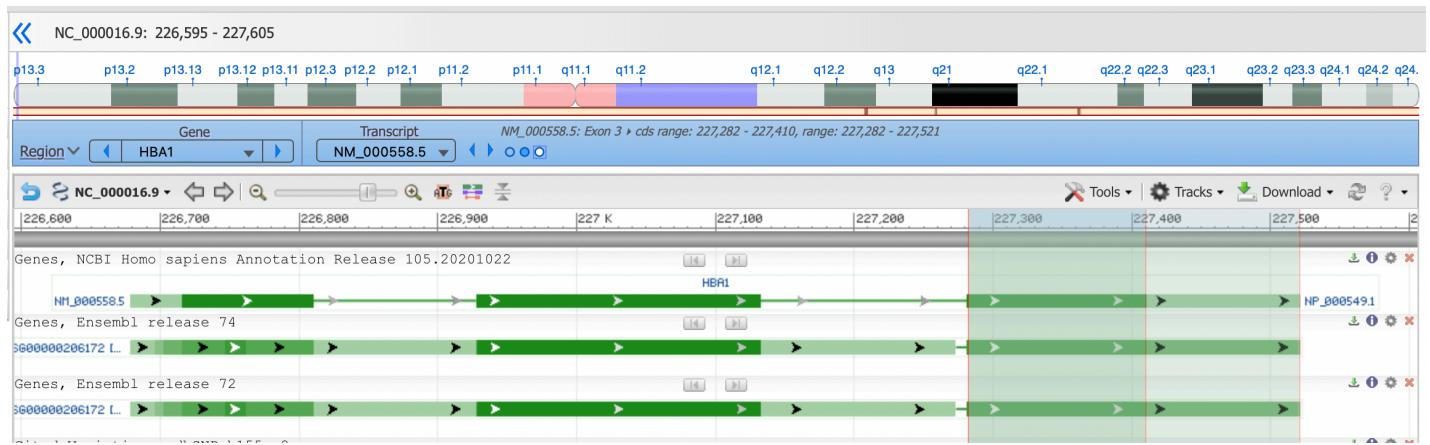
NC_000016.9: 227K..228K (1,011 nt)

Tracks shown: 9/185

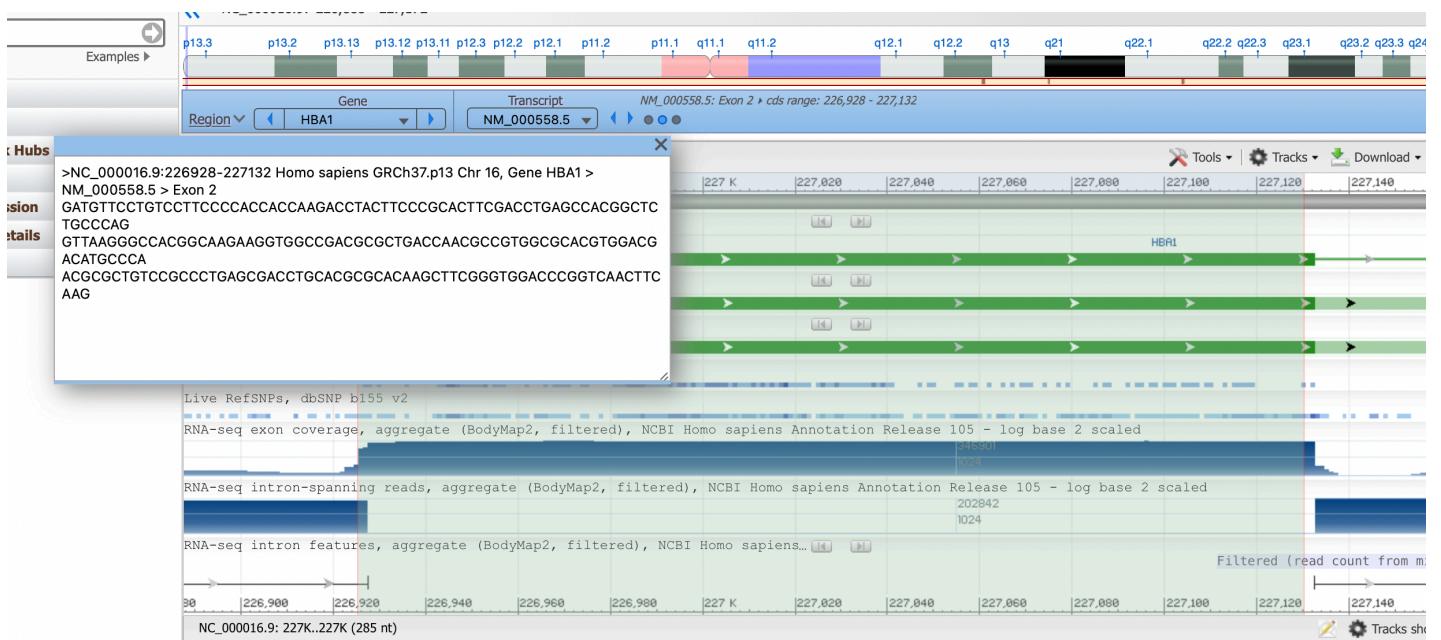
As we may see on the image above, it has an extensive track configuration (for example, we can define an exact set of nucleotides if we want either recommended NCBI track sets (may see such possibilities on the images below) and even create an own NCBI track collection).



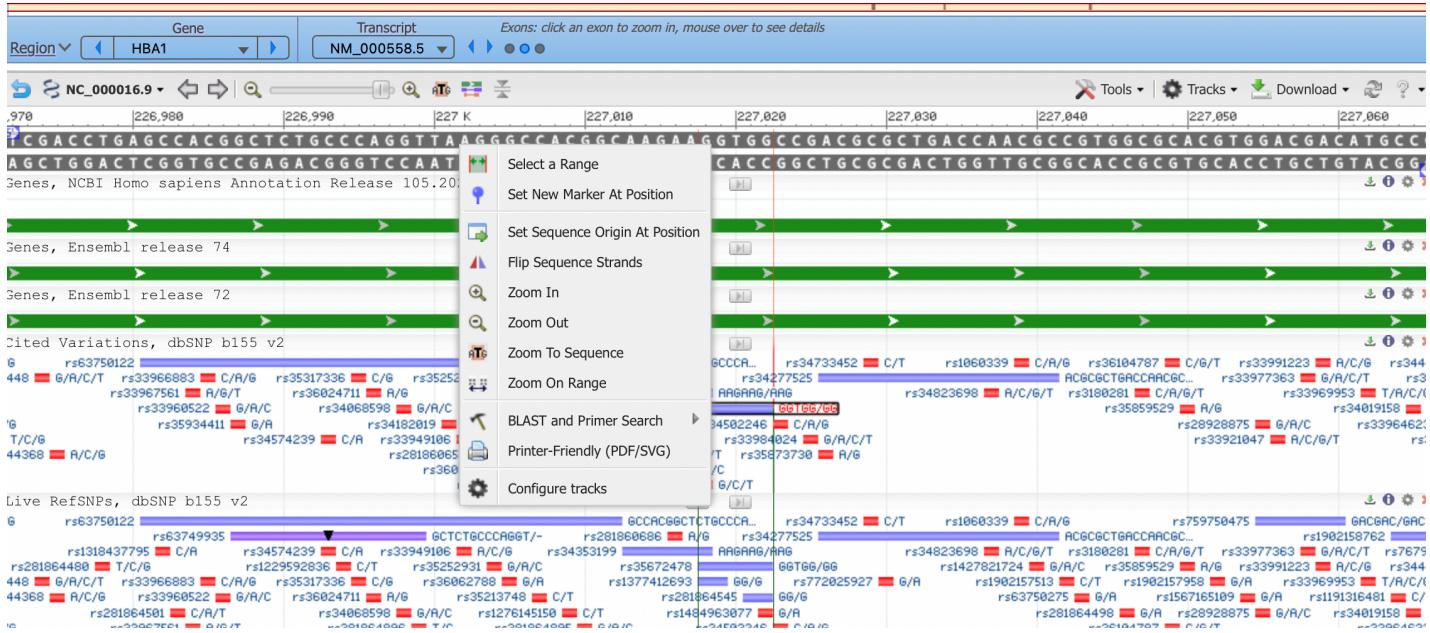
Using a blue bar we may select an exact transcript or an exon location (highlighted in green)



It is possible to view an exact genomic seq if we need:



We may also set a marker at needed position (sometimes it can be useful):



That was just a minor sample of Genome Data Viewer capabilities.

The Genome Data Viewer (GDV) is now the main genome browser at NCBI replacing the Map Viewer, NCBI's original genome browser. GDV is a modern genome browser with essential improvements over Map Viewer. These include sequence-level details and an automated update process that keeps up with the rapid pace of genome sequencing, assembly and annotation.

The GDV home page serves as a main access point for this large collection of organisms and assemblies.

Here are some more advantages that GDV offers over the Map Viewer:

- 1) It is easy to customize the display with tracks representing NCBI's wide range of genomic data.
- 2) It is possible to display own data from a variety of file formats alongside NCBI tracks through the 'Your data' function.
- 3) We can share our displays with others or print publication-quality views.

The NCBI Genome Data Viewer (GDV) allows users to visualize molecular data in a genomic context. The Genome Data Viewer is used by different NCBI resources, such as the Gene Expression Omnibus (GEO) repository, to provide a graphical display of data associated with specified experiments or samples.

The core component of the GDV browser is the NCBI Sequence Viewer, which supports analysis of genomic assemblies at multiple levels, from the whole chromosome or scaffold to the sequence base pair.

The list of default tracks displayed in the Sequence Viewer widget of the GDV is specified by the context of the NCBI application that is using the GDV. For example, when viewing the GDV in the context of GEO, default tracks displayed include: sequence, CpG islands, Genes and dbSNP Cited Variants. It is also good to know that the performance of the GDV may be compromised when loading large (>200) numbers of tracks.

✓ 0 сек. выполнено в 03:40

● ✕