

TUTORIAL GUIDE 1

Part 1: Introduction to UCSC Genome Browser

Understand the functioning and usefulness of a genome browser using the UCSC Genome Browser (<http://genome.ucsc.edu/>)

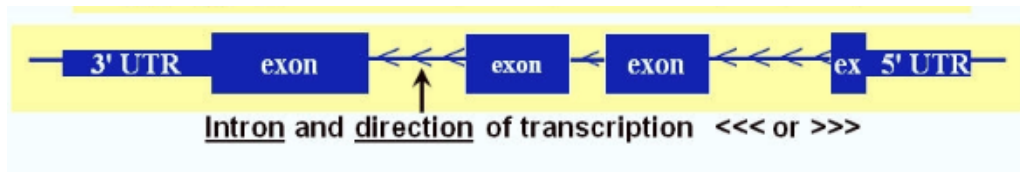
1 - Click on **Genome Browser** on the sidebar or **Genomes** on the top bar.

The screenshot shows the UCSC Genome Browser homepage. At the top, there's a navigation bar with links: Genomes, Genome Browser, Tools, Mirrors, Downloads, My Data, Projects, Help, and About Us. Below this is a large banner image showing a genomic track with various features like genes (BAT2, ATP1B2, TP53, WRAP53) and a heatmap. A search bar is located below the banner with the text "Search genes, data, help docs and more...". Below the search bar, there are two columns: "Tools" and "News". The "Tools" column lists various tools like Genome Browser, BLAT, In-Silico PCR, Table Browser, LiftOver, REST API, and Variant Annotation Integrator. The "News" column lists recent updates and publications. At the bottom, there's a section for "Meetings and Workshops" with links to various conferences.

2 - Search for the TP53 gene (human) using the December 2013 version of the assembly

The screenshot shows the UCSC Genome Browser Gateway search interface. At the top, there's a navigation bar with links: Home, Genomes, Genome Browser, Tools, Mirrors, Downloads, My Data, Projects, Help, and About Us. Below this is a section titled "Browse/Select Species" with a list of popular species: Human, Mouse, Rat, Zebrafish, Fruitfly, Worm, and Yeast. A search bar is located below the species list with the text "Search through thousands of genome browsers" and "Enter species, common name or assembly ID". To the right of the search bar is a section titled "Find Position" with a dropdown menu for "Human Assembly" set to "Dec. 2013 (GRCh38/hg38)". Below this is a text input field for "Position/Search Term" containing "TP53". A "GO" button is located to the right of the input field. Below the input field, the current position is displayed: "chr4:3,074,874-3,074,948". At the bottom, there's a section titled "Human Genome Browser - hg38 assembly" with a "view sequences" button. Below this, there's a section for "UCSC Genome Browser assembly ID: hg38" and "Sequencing/Assembly provider ID: Genome Reference Consortium Human GRCh38.p13 (GCA_000001405.28)". The assembly date is listed as "Dec. 2013 initial release; Dec. 2017 patch release 13".

3 - On the search results page click on the entry "**TP53 (ENST00000269305.9)** at **chr17:7668421-7687490**"



4 - Include 1000 bases upstream (putative promoter) of the chosen transcript:
"chr17:7,667,421-7,687,490"

5 - Try changing the order in which the tracks are displayed in the "viewer"

6 - Change the different levels of display of the Spliced ESTs track

The screenshot displays the UCSC Genome Browser interface for the TP53 gene region on chromosome 17. The browser shows the following tracks and annotations:

- UCSC Genome Browser on Human (GRCh38/hg38)**: The main title of the browser.
- Navigation Tools**: Includes a search bar, a multi-region selection tool, and a zoom in/out tool.
- Genomic Coordinates**: The top track shows the genomic coordinates of the region, ranging from 17p13.1 to 17q25.3.
- Gene Model**: The TP53 gene model is shown below the gene track, with exons represented by blue boxes and introns by lines.
- Annotations**: Various annotations are displayed, including:
 - OMIM Alleles**: A track showing OMIM alleles for the TP53 gene.
 - Gene Expression**: A track showing gene expression data for TP53 across 54 tissues from GTEx RNA-seq of 17382 samples.
 - NCBI RefSeq genes**: A track showing the NCBI RefSeq gene model for TP53.
 - OMIM Gene Phenotypes**: A track showing OMIM gene phenotypes for TP53.

Gencode V48 (Item Details)

Human Gene TP53 (ENST00000269305.9) from GENCODE V48

Description: tumor protein p53, transcript variant 14 (from RefSeq NR_176326.1)
 Gencode Transcript: ENST00000269305.9
 Gencode Gene: ENSG00000141510.19
Coding Region
 Position: hg38 chr17:7,669,609-7,676,594 Size: 6,986 Coding Exon Count: 10

Page Index	Sequence and Links	UniProtKB Comments	Primers	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: 2025-03-26 01:27:12

[-] Sequence and Links to Tools and Databases

Genomic Sequence (chr17:7,668,421-7,687,490)	mRNA (may differ from genome)	Protein (393 aa)
Gene Sorter Genome Browser Other Species FASTA	Gene Interactions Table Schema	AlphaFold
BioGPS Ensembl ExonPrimer	GeneCode GeneCards HGNC	
Lynx MalaCards MGI	neXtProt PubMed Reactome	
UniProtKB Wikipedia		

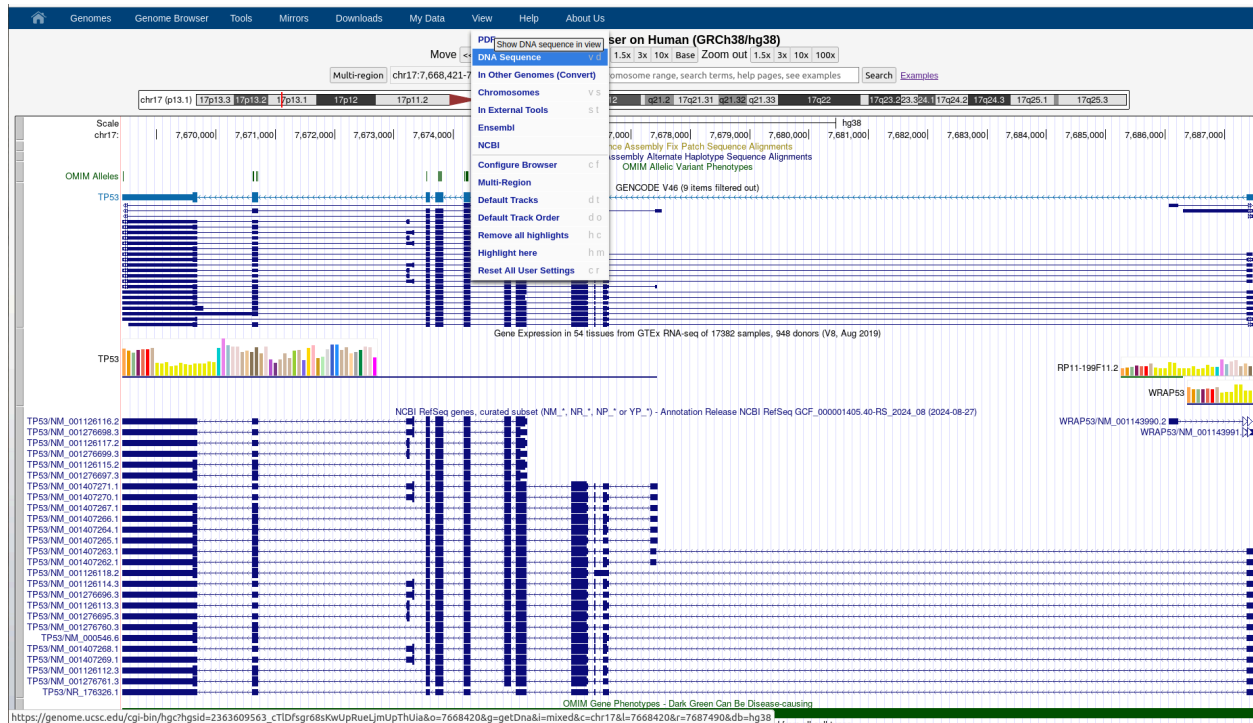
[-] Comments and Description Text from UniProtKB

ID: P53 [HUMAN](#)

DESCRIPTION: RecName: Full=Cellular tumor antigen p53; AltName: Full=Antigen NY-CO-13; AltName: Full=Phosphoprotein p53; AltName: Full=Tumor suppressor p53;
FUNCTION: Acts as a tumor suppressor in many tumor types; induces growth arrest or apoptosis depending on the physiological circumstances and cell type. Involved in cell cycle regulation as a trans-activator that acts to negatively regulate cell division by controlling a checkpoint and seem to have to effect on cell-cycle regulation. Implicated in Notch signaling cross-talk. Prevents CDK7 kinase activity when associated to CAK complex in response to DNA damage, thus stopping cell cycle progression. Isoform 2 enhances the transacti
COFACTOR: Binds 1 zinc ion per subunit.
SUBUNIT: Interacts with AXIN1. Probably part of a complex consisting of TP53, HIPK2 & AXIN1 (By similarity). Binds DNA as a homotrimer. Interacts with histone acetyltransferases EP300 and methyltransferases HRMT1L2 and CARM1, and recruits them to prome
 Found in a complex with CBLESL1 and TP73. Interacts with HIPK1, HIPK2, and TP53BP1. Interacts with WWOX. May interact with HCV core protein. Interacts with USP7 and SYVN1. Interacts with U90AB1. Interacts with CHD8, leading to recruit histone H1 and pr
 ubiquitination and proteasomal degradation of TP53. Directly interacts with FBXO42, leading to ubiquitination and degradation of TP53. Interacts (phosphorylated at Ser-15 by ATM) with the phosphatase PP2A-PPP2R5C holoenzyme; regulates stress-induced TP53-dep
 with PTK2/FYKAL1, this promotes ubiquitination by MDN2. Interacts with PTBKBPYK2, this promotes ubiquitination by MDN2. Interacts with PRKCG. Interacts with human cytomegalovirus/hvH-5 protein UL129.
INTERACTION: Self, NDbExp=3; InAct=EBI-366083, EBI-366083, P03070-, (xeno); NDbExp=6; InAct=EBI-366083, EBI-617698; O15169.AXIN1; NDbExp=4; InAct=EBI-366083, EBI-710484; P10415.BCL2, NDbExp=3; InAct=EBI-366083, EBI-77261; Q07817.1.BCL2L1; NL
 P55060.CSE1L, NDbExp=5; InAct=EBI-366083, EBI-286709; Q14999.CUL7, NDbExp=3; InAct=EBI-366083, EBI-308060; Q8W1T3.CUL9, NDbExp=2; InAct=EBI-366083, EBI-311123; Q9UER7.DAXX, NDbExp=11; InAct=EBI-366083, EBI-77321; Q92841.DDX17, NDbExp=3
 NDbExp=11; InAct=EBI-366083, EBI-1011515; P38646.HSPA9, NDbExp=2; InAct=EBI-366083, EBI-354932; P42858.HTT, NDbExp=4; InAct=EBI-366083, EBI-466029; Q7Z6Z7.HUWE1, NDbExp=3; InAct=EBI-366083, EBI-625934; Q16666-2.F1F16, NDbExp=3; InAct=EBI-3
 EBI-366083, EBI-2689595; P23311.NPXYA, NDbExp=11; InAct=EBI-366083, EBI-389739; Q9V379.NOC2L, NDbExp=8; InAct=EBI-366083, EBI-751547; P06748.NPM1, NDbExp=3; InAct=EBI-366083, EBI-
 P29590.PML, NDbExp=3; InAct=EBI-366083, EBI-295990; Q08752.PPID, NDbExp=4; InAct=EBI-366083, EBI-716596; P36873-1.PPP1CC, NDbExp=2; InAct=EBI-366083, EBI-356289; Q8WUV5.PPP1R1L, NDbExp=3; InAct=EBI-366083, EBI-5550163; Q05397.PTK2, ND
 NDbExp=2; InAct=EBI-366083, EBI-355371; Q9H3D4.TP63, NDbExp=5; InAct=EBI-366083, EBI-2337775; O88988.TP63 (xeno); NDbExp=2; InAct=EBI-366083, EBI-2338025; O15672.TWIST1; NDbExp=6; InAct=EBI-366083, EBI-1797287; Q05086.UBE3A, NDbExp=2; INTA
SUBCELLULAR LOCATION: Cytoplasm. Nucleus. Nucleom, PML body. Endoplasmic reticulum. Note=Interaction with BANP promotes nuclear localization. Recruited into PML bodies together with CHEK2.
 See also [SUBCELLULAR LOCATION](#). Isoform 1: Nucleus. Cytoplasm. Note=Predominantly nuclear but localizes to the cytoplasm when expressed with isoform 4.
SUBCELLULAR LOCATION: Isoform 2: Nucleus. Cytoplasm. Note=Localized mainly in the nucleus with minor staining in the cytoplasm.
SUBCELLULAR LOCATION: Isoform 3: Nucleus. Cytoplasm. Note=Localized in the nucleus in most cells but found in the cytoplasm in some cells.

Close

8 - View the DNA sequence corresponding to the region in the "viewer".



Genomes Genome Browser Tools Mirrors Downloads My Data Projects Help About Us

Get DNA in Window (hg38/Human)

Get DNA for

Position

Note: This page retrieves genomic DNA for a single region. If you would prefer to get DNA for many items in a particular track, or get DNA with formatting options based on gene structure (introns, exons, UTRs, etc.), try using the [Table Browser](#) with the "sequence" output format. You can also use the [REST API](#) with the `/getData/sequence` endpoint function to extract sequence data with coordinates.

Sequence Retrieval Region Options:

Add extra bases upstream (5') and extra downstream (3')

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

Sequence Formatting Options:

☒ All upper case.
☐ All lower case.
☐ Mask repeats: ☒ to lower case ☐ to N
☐ Reverse complement (get '-' strand sequence)

Note: The "Mask repeats" option applies only to "get DNA", not to "extended case/color options".

9 - Starting from **Extended DNA Case/Color Options** choose to display GENCODE V48 in red (254) and Spliced ESTs in green (254). What do the regions colored in yellow represent?

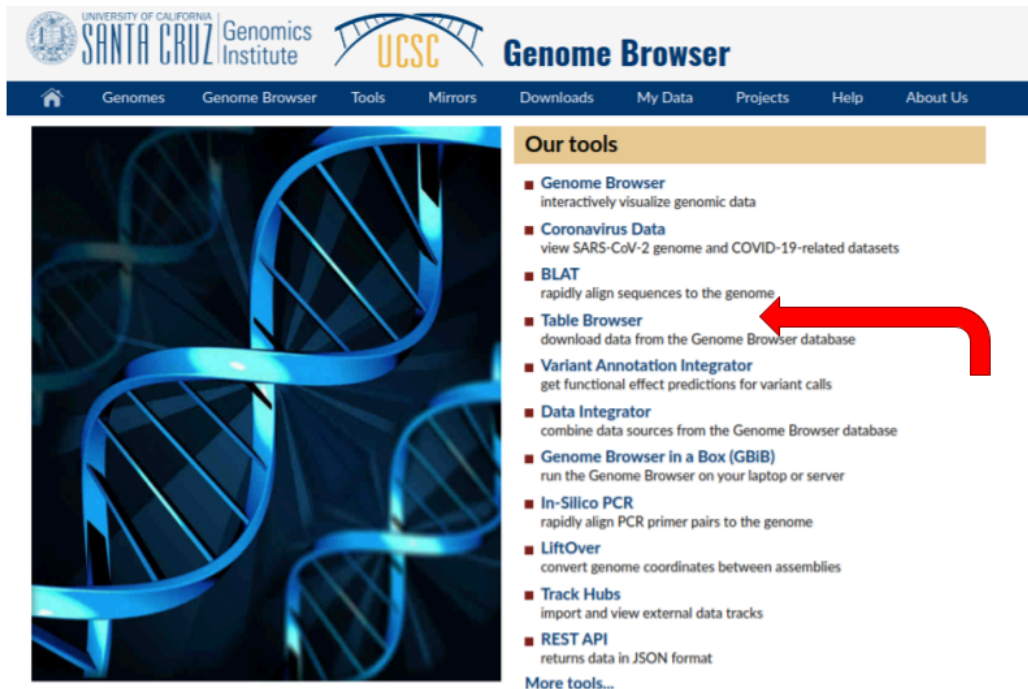
10 - Starting from the page of the details of the chosen isoform (point 7) obtain the protein sequence in FASTA format.

Part 2: Using the UCSC Genome Browser tables

The AIM of this exercise is to understand how to perform advanced searches using the **"table browser"** of the UCSC Genome Browser and to look at the search results with the "custom tracks".

The **"table browser" function** allows you to interact almost directly with the tables of the MySQL database that make up the skeleton of the UCSC GB through some **"custom tracks"**.

1 - Click on Tables Browser in the browser bar



The screenshot shows the UCSC Genome Browser homepage. The header includes the University of California Santa Cruz Genomics Institute logo and the UCSC Genome Browser title. Below the header is a navigation bar with links: Genomes, Genome Browser, Tools, Mirrors, Downloads, My Data, Projects, Help, and About Us. The main content area is divided into two sections. On the left is a large image of a DNA double helix. On the right is a section titled "Our tools" which lists various tools. A red arrow points to the "Table Browser" tool in this list.

Our tools

- **Genome Browser**
interactively visualize genomic data
- **Coronavirus Data**
view SARS-CoV-2 genome and COVID-19-related datasets
- **BLAT**
rapidly align sequences to the genome
- **Table Browser**
download data from the Genome Browser database
- **Variant Annotation Integrator**
get functional effect predictions for variant calls
- **Data Integrator**
combine data sources from the Genome Browser database
- **Genome Browser in a Box (GBIB)**
run the Genome Browser on your laptop or server
- **In-Silico PCR**
rapidly align PCR primer pairs to the genome
- **LiftOver**
convert genome coordinates between assemblies
- **Track Hubs**
import and view external data tracks
- **REST API**
returns data in JSON format

[More tools...](#)

Identify simple repeats with exact CAG sequence in the human genome

1 - Choose the simpleRepeats table using the 2013 assembly of the human genome.

Table Browser

Use this tool to retrieve and export data from the Genome Browser annotation track database. You can limit retrieval based on data attributes and intersect or

Select dataset

clade: Mammal genome: Human assembly: Dec. 2013 (GRCh38/hg38)

group: Repeats track: Simple Repeats

table: simpleRepeat data format description

The simple repeats track contains only one table (simpleRepeats). When there are multiple tables for a track, the main table with genomic location information appears at the top of the table list.

2 - Click on "**summary/statistics**" to get the number of simple repeats present in the human genome

4 - Create a filter to obtain only simple repeats whose sequence is **CAG**

Filter on Fields from hg38.simpleRepeat

bin	is	ignored	0	
chrom	does	match	*	AND
chromStart	is	ignored	0	AND
chromEnd	is	ignored	0	AND
name	does	match	*	AND
period	is	ignored	0	AND
copyNum	is	ignored	0	AND
consensusSize	is	ignored	0	AND
perMatch	is	ignored	0	AND
perIndel	is	ignored	0	AND
score	is	ignored	0	AND
A	is	ignored	0	AND
C	is	ignored	0	AND
G	is	ignored	0	AND
T	is	ignored	0	AND
entropy	is	ignored	0	AND
sequence	does	match	CAG	

AND Free-form SQL query:

Must be a correctly formatted SQL language clause. Here are some Examples:

name like 'ENST%'

name like "ENST*"

name = 'ENST00000693149.1_1'

(name = 'ENST00000693149.1_1' and score < 100) or (name = 'ENST00000691165.1_1' and score < 1000)

Submit Cancel

Identify simple repeats with exact CAG sequence that are found on UCSC genes

- 1 - Click on the **"create" button in the " Intersection with knownGene "** section to reach the intersection creation page.
- 2 - Choose the option **"All Simple Repeats records that have any overlap with GENCODE V48"** and click on **"submit"**
- 3 - Click on **"summary/statistics"** to get the number of simple repeats identified
- 4 - Choose the **"hyperlinks to Genome Browser"** option in the **"output format"** section and click the **"get output" button**
- 5 - Click on the link **"trf at chr1:81501782-81501833"** (Gene ADGRL2)
- 6 - Cliccare su **"trf at chr12:6,936,717-6,936,775"** (Gene ATN1)

Creating Custom Tracks

- 1 - In **"output format"** on the main page of the table browser choose **"custom track"** and then click on **"get output"**
- 2 - Rename the custom track **"SRepeatsGenes"** and change the description to **"Intersection of simple CAG repeats with Genes"**. Finally, click on **"get custom track in genome browser"**

Output simpleRepeat as Custom Track

Custom track header:

name=

description=

visibility=

url=

Create one BED record per:

☒ Whole Gene

☐ Upstream by bases

☐ Downstream by bases

Note: if a feature is close to the beginning or end of a chromosome and upst

- 3 - Move on the gene **"HTT (Homo sapiens huntingtin (HTT), mRNA.)"** and zoom in on the first exon at 5'

4 - Go back to the main page of the table browser and note that custom tracks are available for creating filters and intersections.

5 - Click on "**My Data**" on the top navigation bar and choose the "**custom tracks**" option to display the custom tracks management page.

6 - Connect to the CompGen website

http://compgen.bio.unipd.it/~stefania/Didattica/AA2024-2025/MMOL_BIOINFO_EB/MMOL_BIOINFO_EB.html and download the Practical_session_2.zip file by clicking on "**Guide**" in the line "**II Bioinformatics practical session**"

7 - Unzip the Practical_session_2.zip file and open the BED file with the text editor:

- Information about the default display of our custom track

browser position chr4:56010000-56030000

browser pix 800

browser hide all

browser full knownGene

- Track Features

track name="Items" description="Track for bioinfo2 bioevo" visibility=2 color=0,60,120
useScore=1 db=hg38

- Sequences that will be represented by the "custom track" in BED format

Chr4	56010000	56015000	Item1	100	+
Chr4	56014000	56019000	Item2	200	+
Chr4	56017000	56023000	Item3	800	-
Chr4	56021000	56028000	Item4	300	-

9 - Click on the "**add custom track**" button and paste the custom track on the appropriate field.

clade genome assembly

Display your own data as custom annotation tracks in the browser. Data must be formatted in [bigBed](#), [bigBarChart](#), [bigChain](#), [bigGenePred](#), [bigInteract](#), [bigLolly](#), [bigMaf](#), [bigPsl](#), [bigWig](#), [BAM](#), [barChart](#), [VCF](#), [BED](#), [BED detail](#), [bedGraph](#), [broadPeak](#), [CRAM](#), [GFF](#), [GTF](#), [hic](#), [Interact](#), [MAF](#), [narrowPeak](#), [Personal Genome SNP](#), [PSL](#), or [WIG](#) formats.

- You can paste just the URL to the file, without a "track" line, for bigBed, bigWig, bigGenePred, BAM and VCF.
- To configure the display, set [track](#) and [browser](#) line attributes as described in the [User's Guide](#).

Examples are [here](#). If you do not have web-accessible data storage available, please see the [Hosting](#) section of the Track Hub Help documentation.

Please note a much more efficient way to load data is to use [Track Hubs](#), which are loaded from the [Track Hubs Portal](#) found in the menu under My Data.

Paste URLs or data: Nessun file selezionato

```
browser position chr4:56010000-56030000
browser pix 800
browser hide all
browser pack snp155
browser full knownGene
track name="Items" description="Track per bioinfo2 bioevo" visibility=2
color=0,60,120 useScore=1 db=hg38
```

10 - Click chr4, the default position of our custom track, to view the elements

Summary exercise

Using the tables in the UCSC Genome Browser:

1. Make a custom track via the table browser to represent the subsequences of the isoform "ENST00000269305.9" (GENCODE V48 table) that overlap at least one mRNA (table all_mrna) and view them in the Genome Browser.
2. Obtain the DNA sequence of the custom track and highlight in yellow the sequences that overlap between the GENCODE V48 track and the track of the transcript "ENST00000269305.9". What do yellow sequences represent?