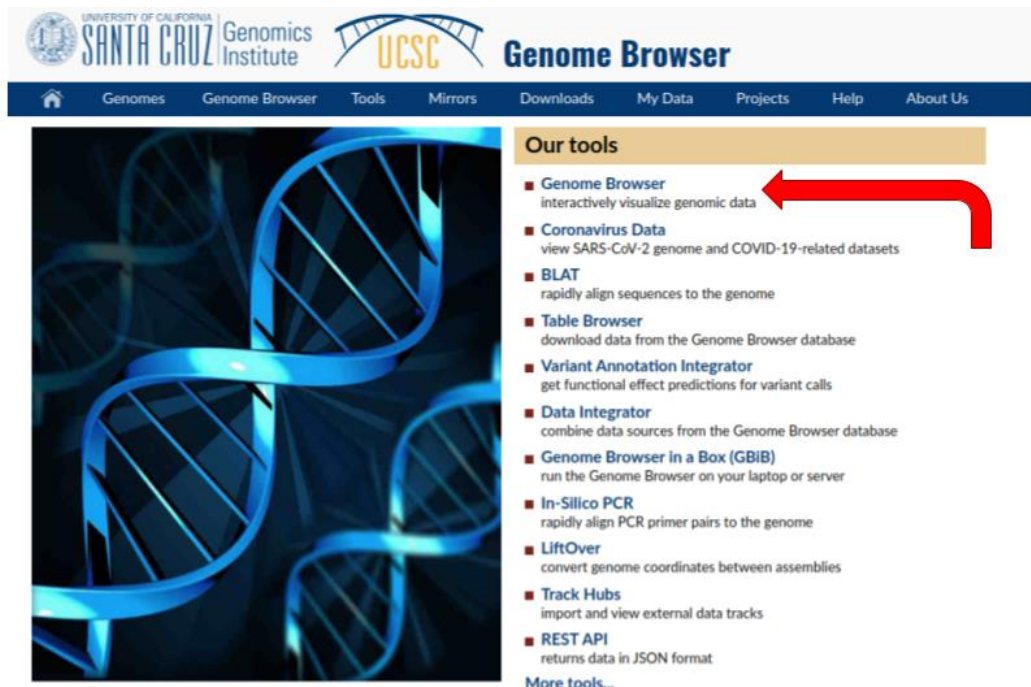


## TUTORIAL GUIDE 1

### 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.



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

UNIVERSITY OF CALIFORNIA  
**SANTA CRUZ** Genomics Institute  
**UCSC** Genome Browser Gateway

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

**Browse/Select Species**

POPULAR SPECIES

Human Mouse Rat Zebrafish Fruitfly Worm Yeast

Search through thousands of genome browsers  
Enter species, common name or assembly ID

Unable to find a genome? Send us a request.

UCSC SPECIES TREE AND CONNECTED ASSEMBLY HUBS

Human Chimp Bonobo

**Find Position**

Human Assembly  
Dec. 2013 (GRCh38/hg38)

Position/Search Term  
TP53

GO

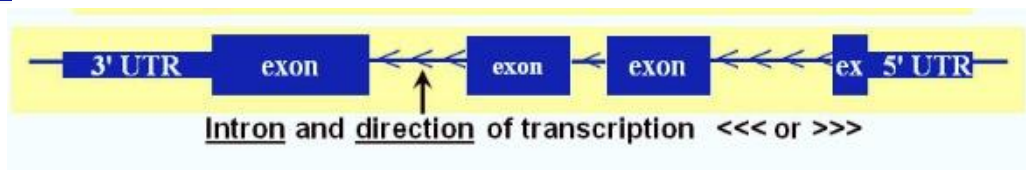
Current position: chr4:3,074,874-3,074,948

Human Genome Browser - hg38 assembly

view sequences

UCSC Genome Browser assembly ID: hg38  
Sequencing/Assembly provider ID: Genome Reference Consortium Human GRCh38.p13 (GCA\_000001405.28)  
Assembly date: 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](#)"

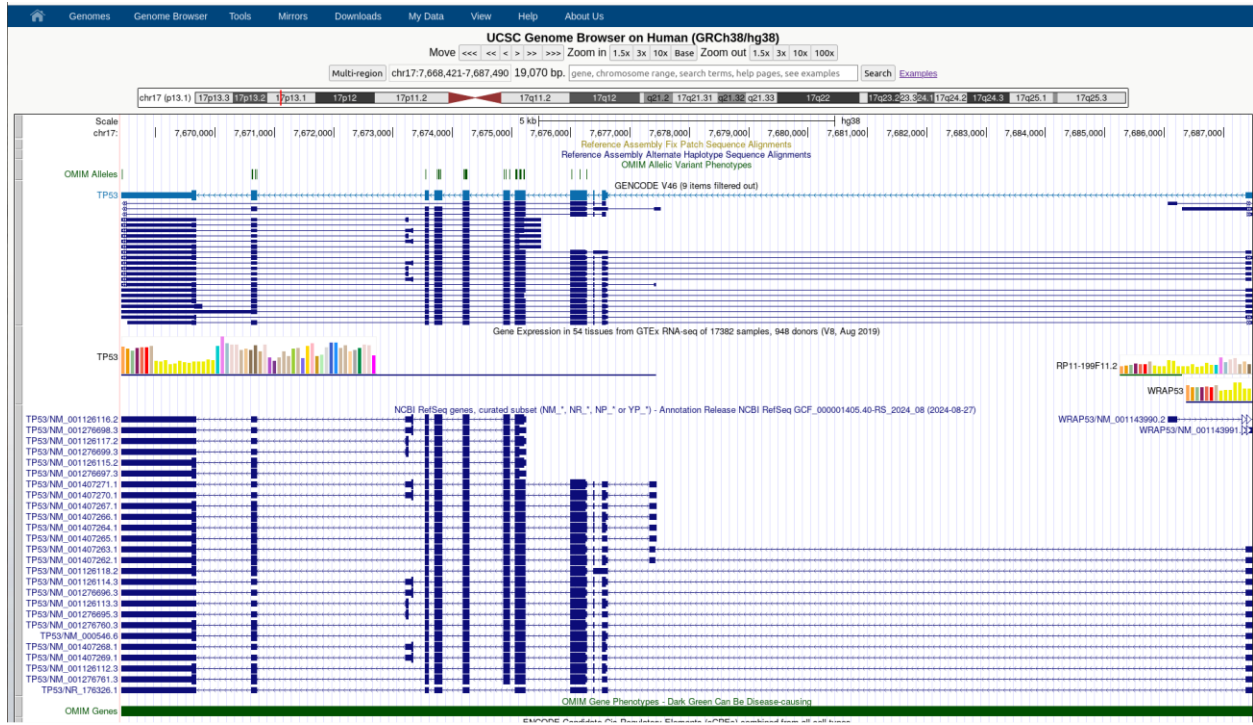


4 - Include 1000 bases upstream (putative promoter) of the chosen transcript: "chr17:7,667,421-7,687,490"

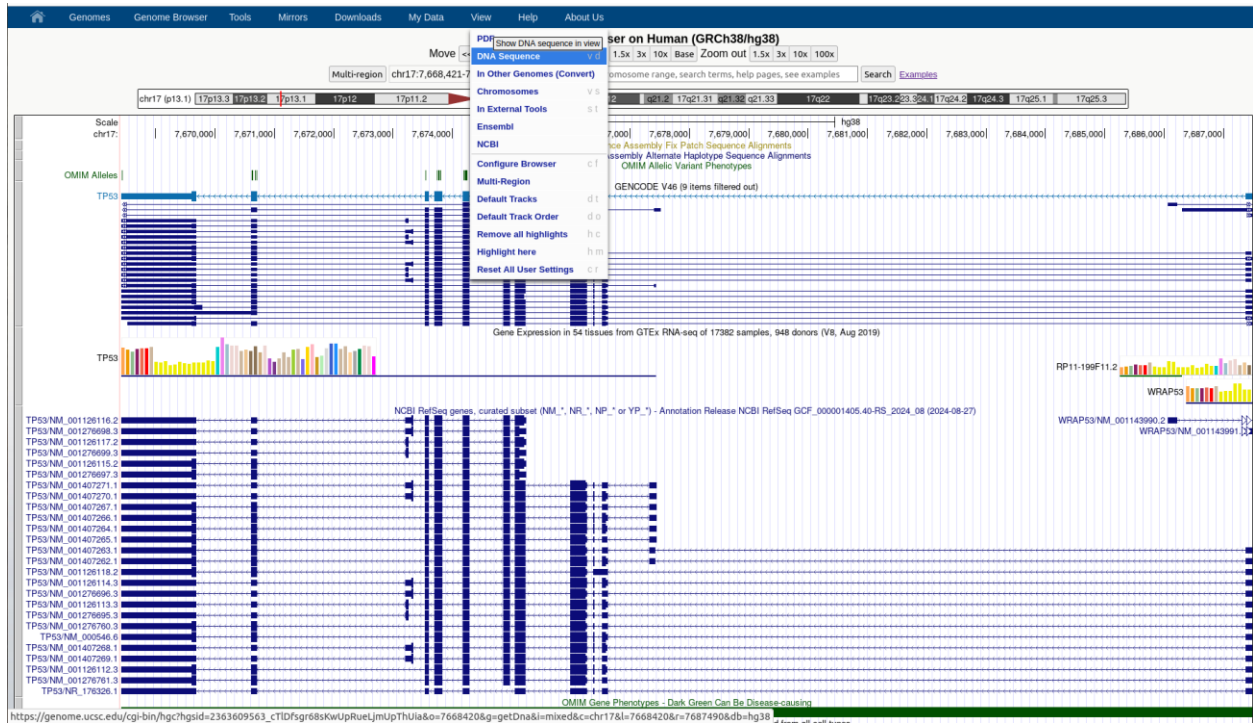
5 - Try changing the order in which the tracks are displayed in the "viewer"

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

7 - View the page on the details of the isoform we have chosen by clicking on the row with the corresponding transcript.



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



**GENCODE V46 (Item Details)**

---

### Human Gene TP53 (ENST00000269305.9) from GENCODE V46

**Description:** Acts as a tumor suppressor in many tumor types; induces growth arrest or apoptosis depending on the physiological circumstances and cell type [PubMed:11025664, PubMed:12524540, PubMed:12810724, PubMed:15186775, PubMed:1534006 PubMed:17317671, PubMed:17349958, PubMed:19556538, PubMed:20673990, PubMed:2099462, PubMed:22726440, PubMed:24051492, PubMed:9840937, PubMed:24652652]. One of the activated genes is an inhibitor of cyclin-dependent kinases. Apopt PIF is involved in activating oxidative stress-induced necrosis; the function is largely independent of transcription. Induces the transcription of long intergenic non-coding RNA p21 (lincRNA-p21) and lincRNA-Mkn1. LincRNA-p21 participates in TP53-depende gene suppression mediated by isoform 1. Isoform 7-mediated apoptosis. Regulates the circadian clock by repressing CLOCK-BMAL1-mediated transcriptional activation of PER2 [PubMed:24051492], (from UniProt P04637)

**PPIF Summary (NM\_000546):** This gene encodes a tumor suppressor protein containing transcriptional activation, DNA binding, and oligomerization domains. The encoded protein responds to diverse cellular stresses to regulate expression of target genes.

codons from identical transcript variants (PMIDs: 12032546, 20937277). [provided by RefSeq, Dec 2016]

**GenCode Transcript:** ENST00000269305.9

**GenCode Gene:** ENSG00000141510.19

**Transcript (including UTRs)**

**Position:** hg38 chr17:7,668,421-7,687,490 **Size:** 19,070 **Total Exon Count:** 11 **Strand:** -

**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	G0 Annotations
mRNA Descriptions	Pathways	Other Names	GeneReviews	Methods	

Data last updated at UCSC: 2024-06-17 15:45:33

---

**- 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
BioGPS	Ensembl	ExonPrimer
Lynx	MalaCards	MG1
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 cc apoptosis and seem to have effect on cell-cycle regulation. Implicated in Notch signaling cross-over. Prevents CDK7 kinase activity when associated to DNA damage, thus stopping cell cycle progression. Isoform 2 enhances the

**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 homotetramer. Interacts with histone acetyltransferases EP300 and methyltransferases HRMT1L2 and CARM1, and recruits them Found in a complex with CBLE1S1 and TP73. Interacts with HIPK1, HIPK2, and TP53INP1. Interacts with WWOX. May interact with HCV core protein. Interacts with USP7 and SYVN1. Interacts with PP59A0B1. Interacts with CHD8, leading to recruit histone f ubiquitination and proteasomal degradation of TP53. Directly interacts with FBXQ42, leading to ubiquitination and degradation of TP53. Interacts (phosphorylated at Ser-15 by ATM) with the phosphatase PP2A-PPP2R5C holoenzyme; regulates stress-induced

**PTK2FAFK1:** this promotes ubiquitination by MDM2. Interacts with PTK2B(PYK2); this promotes ubiquitination by MDM2. Interacts with PRKCG. Interacts with human cytomegalovirusHHV-5 protein UL23.

**INTERACTION:** Self: NBEExp=3; InAct=EBI-366083, EBI-366083; P03070: (gene) NBEExp=6; InAct=EBI-366083, EBI-417698; Q15169 HAI1, NBEExp=4; InAct=EBI-366083, EBI-710484; P10415 BCL2, NBEExp=3; InAct=EBI-366083, EBI-77894; Q07817-1 B PS5060 CSSE11, NBEExp=5; InAct=EBI-366083, EBI-286709; Q14999 CUL7, NBEExp=3; InAct=EBI-366083, EBI-366085; Q8WI73 CULL3, NBEExp=2; InAct=EBI-366083, EBI-331123; Q9URTE DAXX, NBEExp=11; InAct=EBI-366083, EBI-77321; Q92841 DDX3T7, NBEExp=11; InAct=EBI-366083, EBI-1018153; P38646 HSP90A, NBEExp=2; InAct=EBI-366083, EBI-354932; P42858 HTT, NBEExp=4; InAct=EBI-366083, EBI-466029; Q726Z7 HUWE1, NBEExp=3; InAct=EBI-366083, EBI-625934; Q16666-2 IFI16, NBEExp=3; In/ InAct=EBI-366083, EBI-39M373; Q8I7D2 MI-15, NBEExp=4; InAct=EBI-366083, EBI-768959; P23511 NFYA, NBEExp=11; InAct=EBI-366083, EBI-398739; Q9Y739 NOC1, NBEExp=8; InAct=EBI-366083, EBI-751547; P06748 NPML1, NBEExp=3; InAct=EBI-366083,

[Close](#)

## Summary exercise

## Visualize SNPs coding synonymous and missense of mouse NOTCH1 gene

1- Visualize the SNPs (dbSNP version 142) coding synonymous (**Coding - Synonymous**), in the mouse NOTCH1 gene and color them blue

2- We now want to identify the SNPs (dbSNP version 142) **missense variants (Coding - NonSynonymous)** in the mouse NOTCH1 gene and display them in red. Definition of **missense variant**:

*“A genetic alteration in which a single base pair substitution alters the genetic code in a way that produces an amino acid that is different from the usual amino acid at that position. Some missense variants (or mutations) will alter the function of the protein. Also called missense mutation”. From NCI Dictionary of Genetics Terms.*

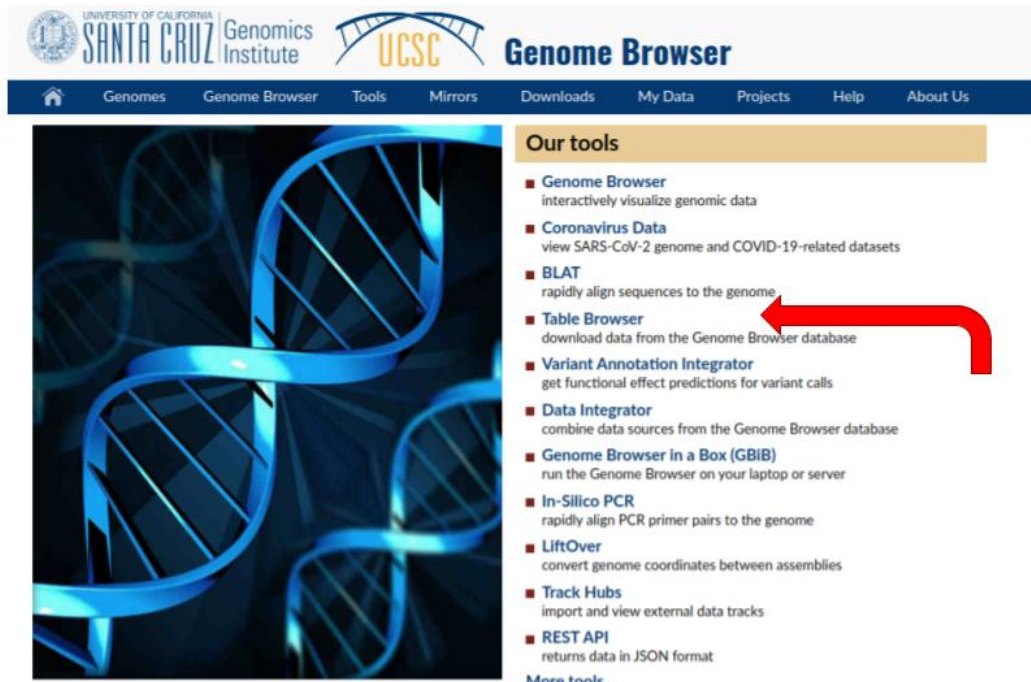
3- "Zoom" this position: chr2:26,467,524-26,467,966 and look at the **Coding annotations by dbSNP section** of the two SNPs found.

## 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



**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:  genome:  assembly:   
group:  track:   
table:

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	<input type="text" value="0"/>	
chrom	does	match	<input type="text" value="*"/>	AND
chromStart	is	ignored	<input type="text" value="0"/>	AND
chromEnd	is	ignored	<input type="text" value="0"/>	AND
name	does	match	<input type="text" value="*"/>	AND
period	is	ignored	<input type="text" value="0"/>	AND
copyNum	is	ignored	<input type="text" value="0"/>	AND
consensusSize	is	ignored	<input type="text" value="0"/>	AND
perMatch	is	ignored	<input type="text" value="0"/>	AND
perIndel	is	ignored	<input type="text" value="0"/>	AND
score	is	ignored	<input type="text" value="0"/>	AND
A	is	ignored	<input type="text" value="0"/>	AND
C	is	ignored	<input type="text" value="0"/>	AND
G	is	ignored	<input type="text" value="0"/>	AND
T	is	ignored	<input type="text" value="0"/>	AND
entropy	is	ignored	<input type="text" value="0"/>	AND
sequence	does	match	<input type="text" value="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)

**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 V46**” 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](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.

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

### Add Custom Tracks

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](#), [bigPsi](#), [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: Or upload:  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:

- Make a custom track via the table browser to represent the subsequences of the isoform "ENST00000269305.9" (GENCODE V46 table) that overlap at least one mRNA (table all\_mrna) and view them in the Genome Browser.
- Obtain the DNA sequence of the custom track and highlight in yellow the sequences that overlap between the GENCODE V46 track and the track of the transcript "ENST00000269305.9". What do yellow sequences represent?