

GUIDA ALL'ESERCITAZIONE 1

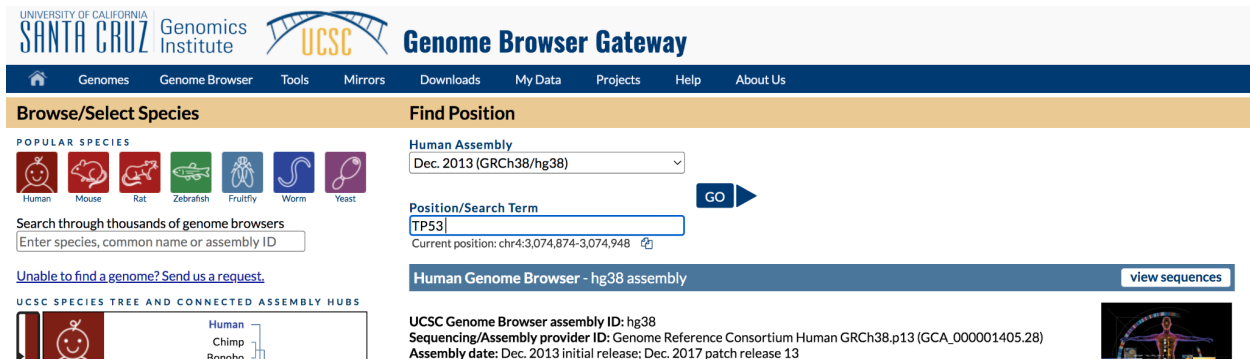
Parte 1: Introduzione all' UCSC Genome Browser

Capire il funzionamento e l'utilità di un genome browser attraverso l'utilizzo dell' UCSC Genome Browser (<http://genome.ucsc.edu/>)

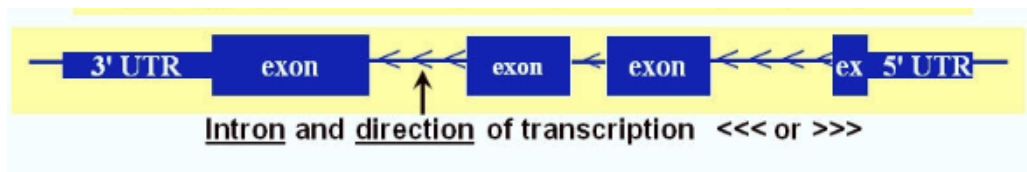
1 - Cliccare su **Genome Browser** nella barra laterale oppure **Genomes** nella barra superiore.



2 - Cercare il gene TP53 (umano) utilizzando la versione di Dicembre 2013 dell'assemblaggio



3 - Nella pagina dei risultati della ricerca cliccare la entry “[TP53 \(ENST00000269305.9\) at chr17:7668421-7687490](#)”



4 - Includere 1000 basi a monte (promotore putativo) del trascritto scelto: “chr17:7,667,421-7,687,490”

5 - Provare a cambiare l'ordine con cui le track vengono visualizzate nel “viewer”

6 - Cambiare i diversi livelli di visualizzazione della track Spliced ESTs

The screenshot displays the UCSC Genome Browser interface for the human genome (GRCh38/hg38). The main title is "UCSC Genome Browser on Human (GRCh38/hg38)". The current view is centered on a specific genomic region on chromosome 17: chr17:7,668,421-7,687,490 (19,070 bp). The interface includes several key components:

- Navigation Bar:** Located at the top, it contains links for Genomes, Genome Browser, Tools, Mirrors, Downloads, My Data, View, Help, and About Us.
- Search and Navigation Controls:** Below the navigation bar, there are controls for moving between regions (multi-region, zoom in/out), a search bar, and a link to examples.
- Track Browser:** On the right side, a track browser lists available tracks, including TP53, RefSeq Curated, OMIM Alleles, Gene Expression, RefSeq genes from NCBI, OMIM Allelic Variant Phenotypes, ENCODE cCREs, Layered H3K27Ac, and 100 vertebrates Basewise Conservation by PhyloP.
- Main Display Area:** This area shows multiple tracks for the selected genomic region. Key tracks include:
 - TP53:** Multiple tracks showing gene models and expression data across different tissues and conditions.
 - RefSeq Curated:** A track showing curated reference sequences, highlighted by a red arrow.
 - OMIM Alleles:** A track displaying clinical significance scores for various alleles.
 - Gene Expression:** A heatmap showing gene expression levels across 54 tissues from GTEx RNA-seq data.
 - RefSeq genes from NCBI:** A track listing genes from the NCBI Reference Sequence database.
 - OMIM Allelic Variant Phenotypes:** A track showing phenotypic associations for OMIM variants.
 - ENCODE cCREs:** A track highlighting candidate cis-regulatory elements (cCREs) from the ENCODE project.
 - Layered H3K27Ac:** A track showing histone marks (H3K27Ac) across different cell lines.
 - 100 vertebrates Basewise Conservation by PhyloP:** A track illustrating conservation across 100 vertebrate species.

The image shows the UCSC Genome Browser interface for the TP53 gene on chromosome 17. The main track displays the gene structure with exons as blue boxes and introns as lines with arrows. Below the gene structure are various tracks including RefSeq genes from NCBI, OMIM Alleles, Gene Expression in 54 tissues from GTEx, ENCODE eCREs, Layered H3K27Ac, and 100 vertebrates Basewise Conservation by PhyloP. The right sidebar contains navigation and configuration options like 'Configure Browser', 'Multi-Region', 'Default Tracks', and 'Reset All User Settings'.

GenomesGenome BrowserToolsMirrorsDownloadsMy DataProjectsHelpAbout Us

Get DNA in Window (hg38/Human)

Get DNA for

Position chr17:7,668,421-7,687,490

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 - A partire da **Extended DNA Case/Color Options** scegliere di visualizzare GENCODE V41 in rosso (254) e Spliced ESTs in verde (254). Cosa rappresentano le regioni colorate in giallo?

10 - A partire dalla pagina sui dettagli dell'isoforma scelta (punto 7) ottenere la sequenza proteica in formato FASTA.

GenomesGenome BrowserToolsMirrorsDownloadsMy DataProjectsHelpAbout Us

Human Gene TP53 (ENST00000269305.9) from GENCODE V41

Description: Homo sapiens tumor protein p53 (TP53), transcript variant 1, mRNA. (from RefSeq NM_000546)

RefSeq Summary (NM_000546): This gene encodes a tumor suppressor protein containing transcriptional activation, DNA binding, and oligomerization domains. The encoded protein responds to diverse initiation codons from identical transcript variants (PMIDs: 12032546, 20937277). [provided by RefSeq, Dec 2016].

Gencode Transcript: ENST00000269305.9

Gencode Gene: ENSG00000141510.18

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	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: 2022-05-14 18:57:26

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
CGAP	Ensembl	Entrez Gene
HGNC	HPRD	Lynx
PubMed	Reactome	UniProtKB

Gene interactions	Table Schema	BioGPS
ExonPrimer	Gencode	GeneCards
MGI	neXtProt	OMIM
Wikipedia		

Esercizio di riepilogo

Visualizzare gli SNPs coding synonymous e missense del gene notch1 del topo

1- Visualizzare gli SNPs (dbSNP versione 142) coding synonymous (**Coding - Synonymous**), nel gene NOTCH1 di topo e colorarli di blu

2- Vogliamo ora identificare gli SNPs (dbSNP versione 142) **missense variants (Coding - NonSynonymous)** nel gene NOTCH1 di topo e visualizzarli in rosso. Definizione di **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- “Zoommare” questa posizione: chr2:26,467,082-26,468,407 e dare un’occhiata alla sezione **Coding annotations by dbSNP** dei due SNP trovati.

Parte 2: Uso dell'UCSC Genome Browser tables

L'obiettivo dell'esercitazione è capire come effettuare ricerche avanzate utilizzando il “**table browser**” dell'UCSC Genome Browser e visualizzare i risultati delle ricerche con le “custom tracks”.

La funzione “**table browser**” permette di interagire in maniera quasi diretta con le tabelle del database MySQL che costituiscono lo scheletro dell'UCSC GB tramite delle “**custom track**”.

1 - Cliccare su Tables Browser nella barra di navigazione



Identificare nel genoma umano le simple repeats con sequenza esatta CAG

1 - Scegliere la tabella simpleRepeats utilizzando l'assemblaggio del 2013 del genoma umano.

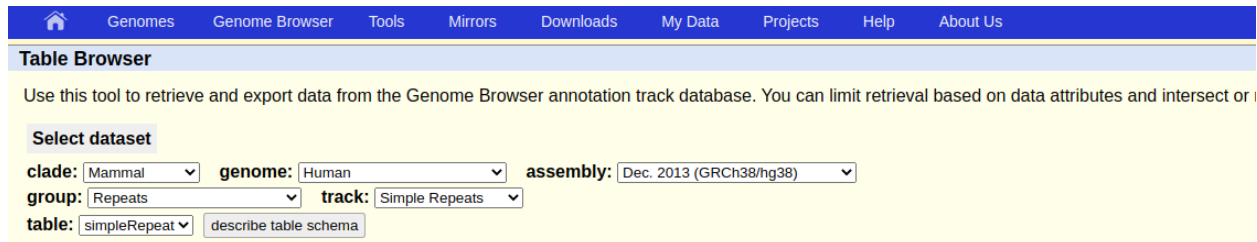


Table Browser

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:

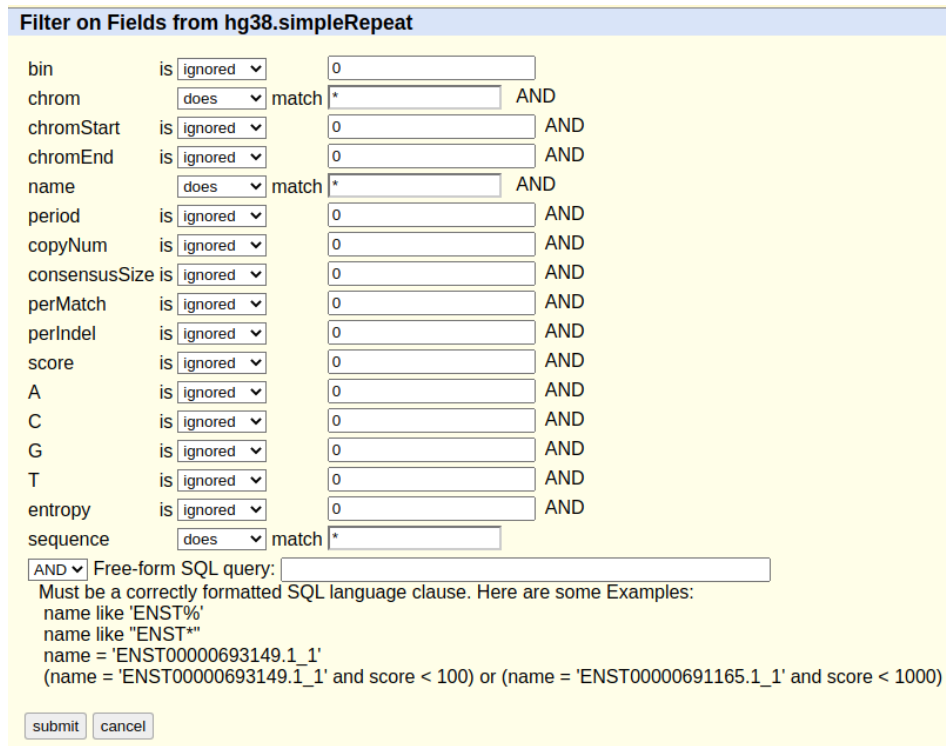
table: [describe table schema](#)

La track simple repeats contiene solo una tabella (simpleRepeats). Quando ci sono più tabelle per una track, la tabella principale con informazioni di posizione genomica appare in prima posizione nella lista delle tabelle.

2 - Cliccare sul tasto “**describe table schema**” per vedere una descrizione della tabella

3 - Cliccare su “**summary/statistics**” per ottenere il numero di simple repeats presenti nel genoma umano

4 - Creare un filtro per ottenere soltanto simple repeats la cui sequenza sia **CAG**



Filter on Fields from hg38.simpleRepeat

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

AND

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)

Identificare simple repeats con sequenza esatta CAG che si trovano su geni UCSC

- 1 - Cliccare sul tasto **“create”** nella sezione **“intersection”** per raggiungere la pagina di creazione delle intersezioni.
- 2 - Scegliere l'opzione **“All Simple Repeats records that have any overlap with GENCODE V41”** e cliccare sul tasto **“submit”**
- 3 - Cliccare su **“summary/statistics”** per ottenere il numero di simple repeats identificate
- 4 - Scegliere l'opzione **“hyperlinks to Genome Browser”** nella sezione **“output format”** e cliccare il tasto **“get output”**
- 5 - Cliccare sul link **“trf at chr1:81501782-81501833”** (Gene ADGRL2)
- 6 - Cliccare su **“trf at chr12:6,936,717-6,936,775”** (Gene ATN1)

Creazione di custom tracks

- 1 - In **“output format”** nella pagina principale del table browser scegliere **“custom track”** e poi cliccare su **“get output”**
- 2 - Rinominare **“SRepeatsGenes”** la custom track e cambiare la descrizione a **“Intersection of simple CAG repeats with Genes”**. Infine, cliccare su **“get custom track in genome browser”**

The screenshot shows a web form titled "Output simpleRepeat as Custom Track". It contains the following fields and options:

- Custom track header:**
 - name=** SRepeatsGenes
 - description=** Intersection of simple CAG repeats with Genes
 - visibility=** pack (dropdown menu)
 - url=** (empty text box)
- Create one BED record per:**
 - ☒ Whole Gene
 - ☐ Upstream by 200 bases
 - ☐ Downstream by 200 bases
- Note:** if a feature is close to the beginning or end of a chromosome and upst
- Buttons at the bottom: "get custom track in table browser", "get custom track in file", "get custom track in genome browser", and "cancel".

- 3 - Spostarsi sul gene **“HTT (Homo sapiens huntingtin (HTT), mRNA.)”** e zoomare sul primo esone al 5’

4 - Tornare alla pagina principale del table browser e notare che le custom track sono disponibili per la creazione di filtri e intersezioni.

5 - Cliccare su **“My Data”** sulla barra di navigazione superiore e scegliere l’opzione **“custom tracks”** per visualizzare la pagina di gestione delle custom tracks.

6 - Collegarsi al sito di CompGen

http://compgen.bio.unipd.it/~stefania/Didattica/AA2022-2023/MMOL_BIOINFO_BE/MMOL_BIOINFO_BE.html e scaricare il file esercitazione1.zip cliccando su **“Guida”** della riga **“I esercitazione BIOINFORMATICA”**

7 - Decomprimere il file esercitazione1.zip e aprire il file BED con l’editor di testo:

- Informazioni riguardanti il display di default della nostra custom track

browser position chr4:56010000-56030000

browser pix 800

browser hide all

browser full knownGene

- Caratteristiche della track

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

- Sequenze che verranno rappresentate dalla “custom track” in formato BED

chr4	56010000	56015000	Item1	100	+
chr4	56014000	56019000	Item2	200	+
chr4	56017000	56023000	Item3	800	-
chr4	56021000	56028000	Item4	300	-

9 - Cliccare sul tasto **“add custom track”** e incollare la custom track sull’apposito campo.

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

Add Custom Tracks

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

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: Or upload: Scogli file Nessun file selezionato Submit

```
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
```

Clear

10 - Cliccare chr4, la posizione di default della nostra custom track, per visualizzare gli elementi

Esercizio di riepilogo

Utilizzando le tabelle dell'UCSC Genome Browser:

1. Fare una custom track tramite il table browser per rappresentare le sottosequenze dell'isoforma "ENST00000269305.9" (tabella GENCODE V41) che si sovrappongono ad almeno un mRNA (tabella all_mrna) e vederle nel Genome Browser.
2. Ottenere la sequenza di DNA della custom track ed evidenziare in giallo le sequenze che si sovrappongono tra UCSC gene track e la track del trascritto "ENST00000269305.9". Cosa rappresentano le sequenze gialle?