

GUIDA ALL'ESERCITAZIONE 1

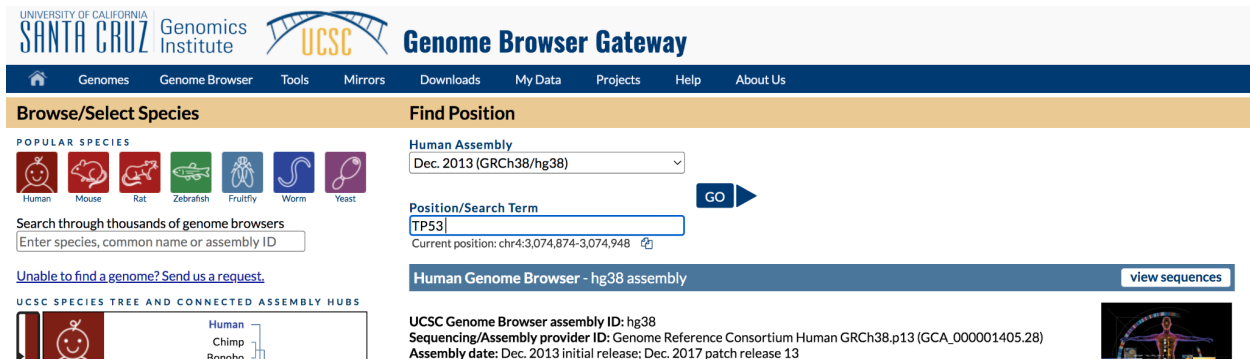
Parte 1: Introduzione all' UCSC Genome Browser

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

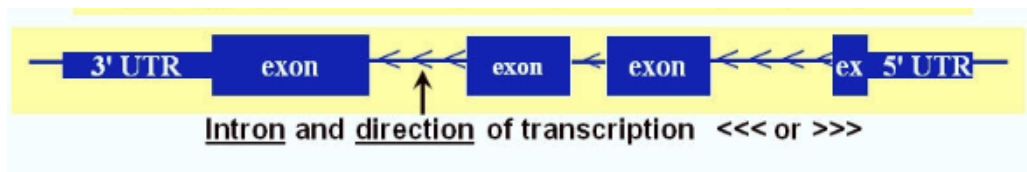
1 - Cliccare su **Genome Browser** nella barra laterale oppure **Genomes** nella barra superiore.



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](#)”

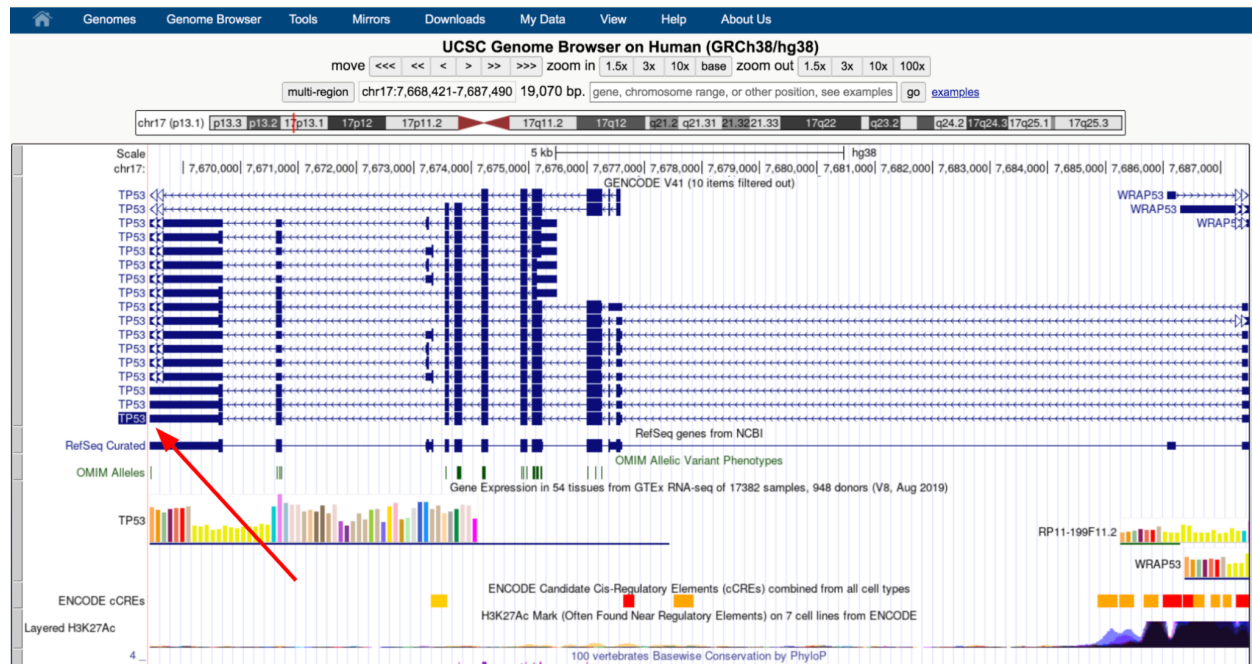


4 - Includere 1000 basi a monte (promotore putativo) del trascritto scelto: “chr17:7,667,421-7,687,490”

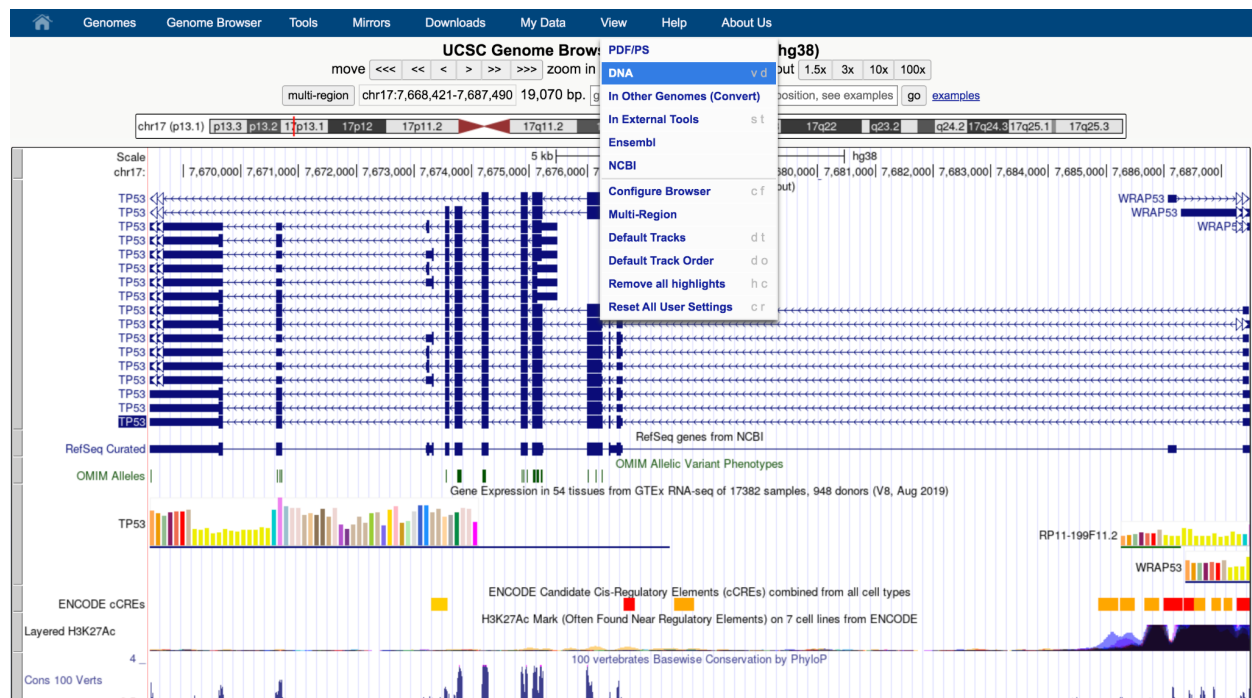
5 - Provare a cambiare l'ordine con cui le track vengono visualizzate nel “viewer”

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

7 - Visualizzare la pagina sui dettagli dell'isoforma da noi scelta cliccando sulla riga con il corrispondente trascritto.



8 - Visualizzare la sequenza di DNA corrispondente alla regione nel “viewer”.



[Home](#)
[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 - A partire da **Extended DNA Case/Color Options** scegliere di visualizzare GENCODE V44 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.

[Home](#)
[Genomes](#)
[Genome Browser](#)
[Tools](#)
[Mirrors](#)
[Downloads](#)
[My Data](#)
[Projects](#)
[Help](#)
[About 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	Gene interactions	Table Schema	BioGPS
CGAP	Ensembl	Entrez Gene	ExonPrimer	Gencode	GeneCards
HGNC	HPRD	Lynx	MGI	neXtProt	OMIM
PubMed	Reactome	UniProtKB	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: chr2:26,467,524-26,467,966 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: Mammal genome: Human assembly: Dec. 2013 (GRCh38/hg38)

group: Repeats track: Simple Repeats

table: simpleRepeat data format description

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 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	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	*	

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

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 V44”** 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”**

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 - 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/AA2023-2024/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.

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 V44) 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 GENCODE V44 track e la track del trascritto "ENST00000269305.9". Cosa rappresentano le sequenze gialle?