

# Bioinformatica II

LM Biologia Evoluzionistica, Università di Padova

Docente: Dr. Stefania Bortoluzzi

## Esercitazione 2

Padova, 26 ottobre 2016

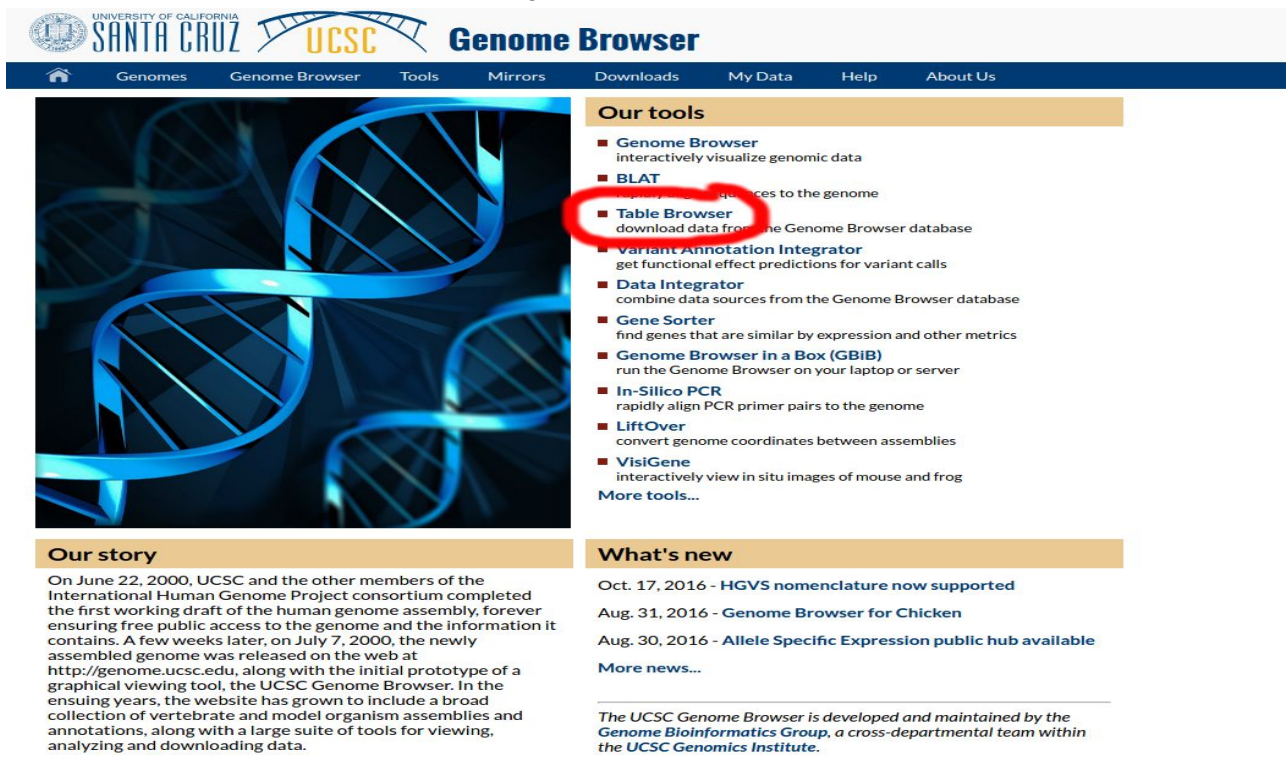
### GUIDA

#### *Uso dell'UCSC Genome Browser tables*

#### Obiettivo dell'esercitazione

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.

1 - Cliccare su Tables nella barra di navigazione superiore



**UCSC Genome Browser**

Genomes Genome Browser Tools Mirrors Downloads My Data Help About Us

**Our tools**

- **Genome Browser**  
interactively visualize genomic data
- **BLAT**  
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
- **Gene Sorter**  
find genes that are similar by expression and other metrics
- **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
- **VisiGene**  
interactively view in situ images of mouse and frog

[More tools...](#)

**Our story**

On June 22, 2000, UCSC and the other members of the International Human Genome Project consortium completed the first working draft of the human genome assembly, forever ensuring free public access to the genome and the information it contains. A few weeks later, on July 7, 2000, the newly assembled genome was released on the web at <http://genome.ucsc.edu>, along with the initial prototype of a graphical viewing tool, the UCSC Genome Browser. In the ensuing years, the website has grown to include a broad collection of vertebrate and model organism assemblies and annotations, along with a large suite of tools for viewing, analyzing and downloading data.

**What's new**

- Oct. 17, 2016 - **HGVS nomenclature now supported**
- Aug. 31, 2016 - **Genome Browser for Chicken**
- Aug. 30, 2016 - **Allele Specific Expression public hub available**

[More news...](#)

The UCSC Genome Browser is developed and maintained by the **Genome Bioinformatics Group**, a cross-departmental team within the **UCSC Genomics Institute**.

Il Genome browser è composto da tabelle di dati, alcune di queste tabelle sono primarie e contengono informazione di posizione, per esempio la tabella degli UCSC Genes. Ci sono inoltre tabelle ausiliari che possono anche non avere informazioni di posizione.

## Identificare nel genoma umano le simple repeats con sequenza esatta CAG

1 - Scegliere la tabella simpleRepeats utilizzando l'assemblaggio del 2009 del genoma umano.

**Table Browser**

Use this program to retrieve the data associated with a track in text format, to calculate intersections between tracks, and to retrieve DNA sequence covered by a track. For help in using this application see [Using the Table Browser](#) for a description of the controls in this form, the [User's Guide](#) for general information and sample queries, and the OpenHelix Table Browser [tutorial](#) for a narrated presentation of the software features and usage. For more complex queries, you may want to use [Galaxy](#) or our [public MySQL server](#). To examine the biological function of your set through annotation enrichments, send the data to [GREAT](#). Send data to [GenomeSpace](#) for use with diverse computational tools. Refer to the [Credits](#) page for the list of contributors and usage restrictions associated with these data. All tables can be downloaded in their entirety from the [Sequence and Annotation Downloads](#) page.

clade:  genome:  assembly:

group:  track:

table:

region: ☒ genome ☐ ENCODE Pilot regions ☐ position

identifiers (names/accessions):

filter:

intersection:

correlation:

output format:  Send output to ☐ [Galaxy](#) ☐ [GREAT](#) ☐ [GenomeSpace](#)

output file:  (leave blank to keep output in browser)

file type returned: ☒ plain text ☐ gzip compressed

To reset all user cart settings (including custom tracks), [click here](#).

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 hg19.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 query:

5 - Selezionare l'opzione "all fields from selected table" nel campo "output format" e cliccare il tasto "get output".

### 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 UCSC Genes" 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 chr4:3076604-3076667](#)" (Gene HTT)

6 - Cliccare su [trf at chr12:7045880-7045938](#) (Gene ATN11)

7 - Cliccare su [trf at chr17:17697094-17697134](#) (Gene RAI1)

8 - Cliccare su [trf at chr19:46273463-46273524](#) (DMPK)

## 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 - Creare un file di testo contenente una “custom track” in formato BED:

chr4	56010000	56015000	Item1	900	+
chr4	56014000	56019000	Item2	200	+
chr4	56017000	56023000	Item3	800	-
chr4	56021000	56028000	Item4	300	-

7 - Inserire nella custom track le informazioni riguardanti il display di default della nostra custom track

browser position chr4:56010000-56030000  
browser pix 800

browser hide all  
browser pack snp130  
browser full knownGene

8 - Inserire le caratteristiche della track

track name="Items" description="Track per bioinfo2 bioevo" visibility=2 color=200,0,200  
useScore=1 db=hg18

9 - Cliccare sul tasto “add custom track” e incollare la custom track sull’apposito campo.  
Cambiare l’assembly del genoma a hg18 e cliccare submit.

**Add Custom Tracks**

clade  genome  assembly

Display your own data as custom annotation tracks in the browser. Data must be formatted in [BED](#), [bigBed](#), [bedGraph](#), [broadPeak](#), [narrowPeak](#), or [PSL](#) formats. To configure the display, set [track](#) and [browser](#) line attributes as described provided via only a URL or embedded in a track line in the box below. Publicly available custom tracks are listed [here](#).

Paste URLs or data: Or upload:  No file chosen

```
browser position chr4:56010000-56030000  
browser pix 800  
browser hide all  
browser pack snp130  
browser full knownGene  
track name="Items" description="Track per bioinfo2 bioevo" visibility=2  
color=200,0,200 useScore=1 db=hg18
```

Optional track documentation: Or upload:  No file chosen

Click [here](#) for an HTML document template that may be used for Genome Browser track descriptions.

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

## Esercizio

Utilizzando le tabelle dell'UCSC Genome Browser, ottenere le sottosequenze del trascritto uc002gij.3 (tabella UCSC Gene) che si sovrappongono ad almeno un mRNA (tabella all\_mrna).