

# Bioinformatica II

LM Biologia Evoluzionistica, Università di Padova

Docente: Prof. Stefania Bortoluzzi

## Esercitazione 1

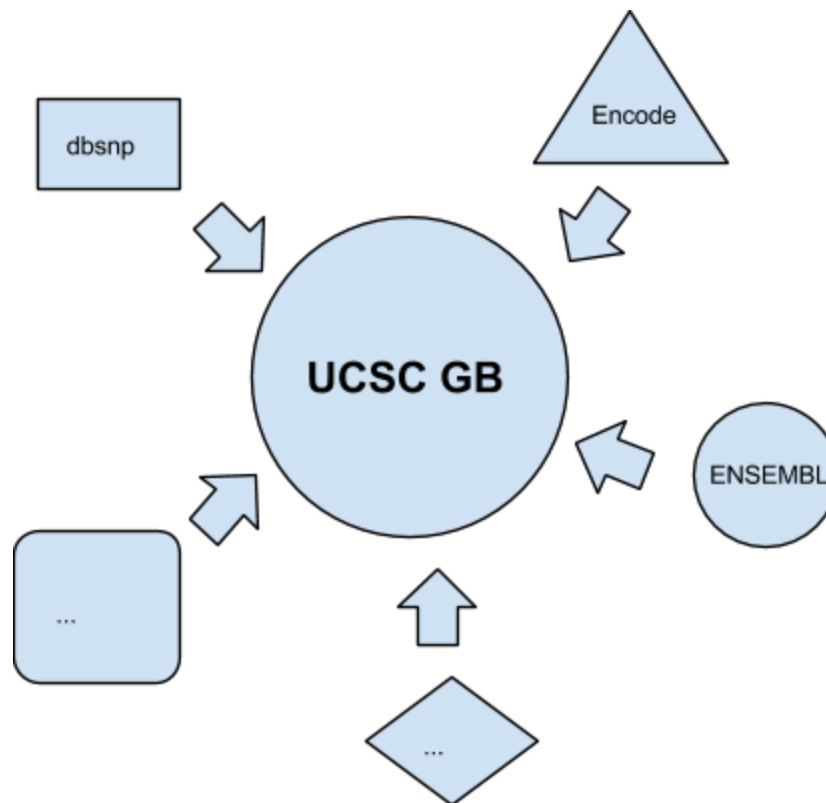
Padova, 19 ottobre 2016

### GUIDA

#### *Introduzione all' UCSC Genome Browser*

##### Obbiettivo dell'esercitazione

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



## Ricerche di base

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

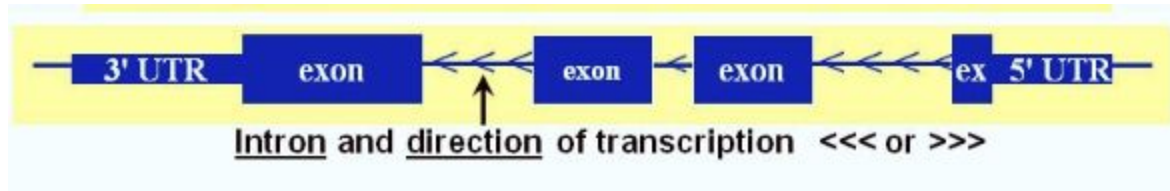
The screenshot shows the UCSC Genome Browser homepage. The top navigation bar includes links for Genomes, Genome Browser, Tools, Mirrors, Downloads, My Data, Help, and About Us. The main content area features a large DNA double helix image on the left and a list of tools on the right. A red arrow points to the 'Genome Browser' tool, which is described as 'interactively visualize genomic data'. Other tools listed include BLAT, Table Browser, Variant Annotation Integrator, Data Integrator, Gene Sorter, Genome Browser in a Box (GBIB), In-Silico PCR, LiftOver, and VisiGene. The 'Our story' and 'What's new' sections are also visible at the bottom.

2 - Cercare il gene tp53 (umano) utilizzando la versione di febbraio 2009 dell'assemblaggio

The screenshot shows the UCSC Genome Browser Gateway search interface. The 'Browse/Select Species' section is highlighted with a red circle, showing a list of species including Human, Mouse, Rat, Fruitfly, Worm, and Yeast. The 'Find Position' section is also highlighted with a red circle, showing a search for 'Human Assembly' (Feb. 2009 GRCh37/hg19) and a 'Position/Search Term' of 'tp53'. The 'GO' button is also highlighted. Below the search results, the 'Human Genome Browser - hg19 assembly' section is visible, providing information about the February 2009 human reference sequence (GRCh37) and sample position queries.

Request:	Genome Browser Response:
chr7	Displays all of chromosome 7
chrUn_gi000212	Displays all of the unplaced contig gi000212
20p13	Displays region for band p13 on chr 20
chr3:1-1000000	Displays first million bases of chr 3, counting from p-arm telomere
chr3:1000000+2000	Displays a region of chr3 that spans 2000 bases, starting with position 1000000

3 - Nella pagina dei risultati della ricerca cliccare la entry “TP53 (uc002gij.3) at chr17:7571720-7590868

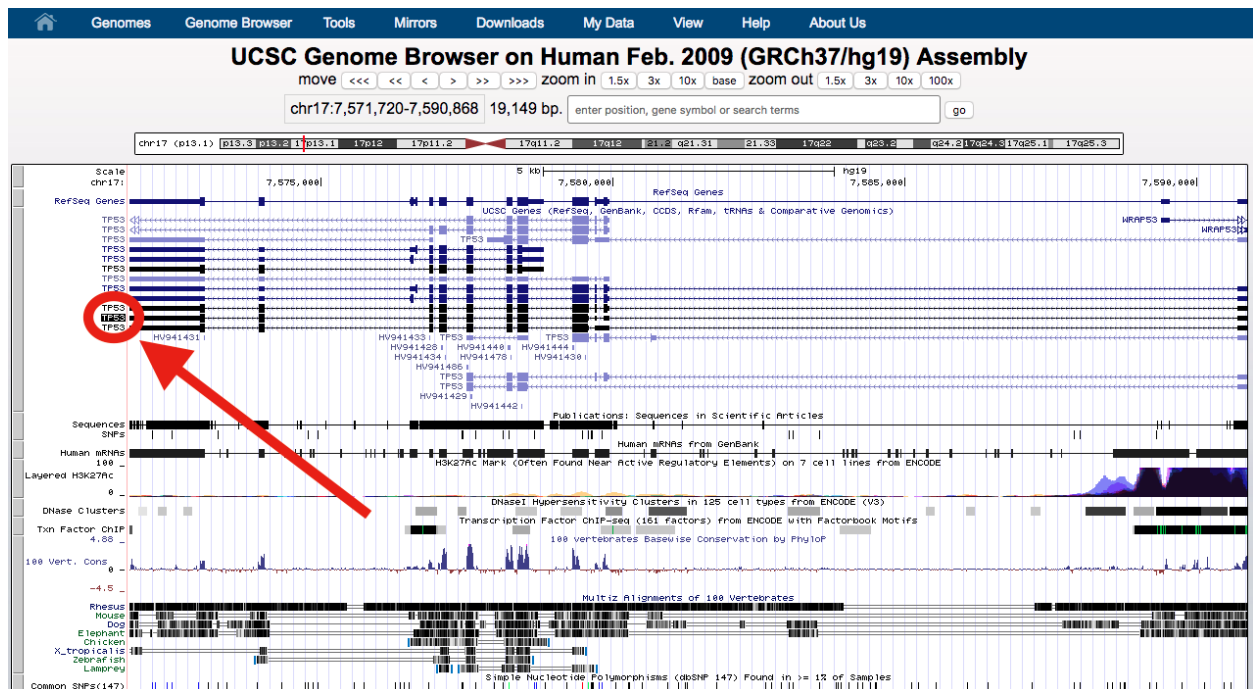


4 - Includere 1000 basi a valle (promotore putativo) del trascritto scelto: “chr17:7,571,720-7,591,868”

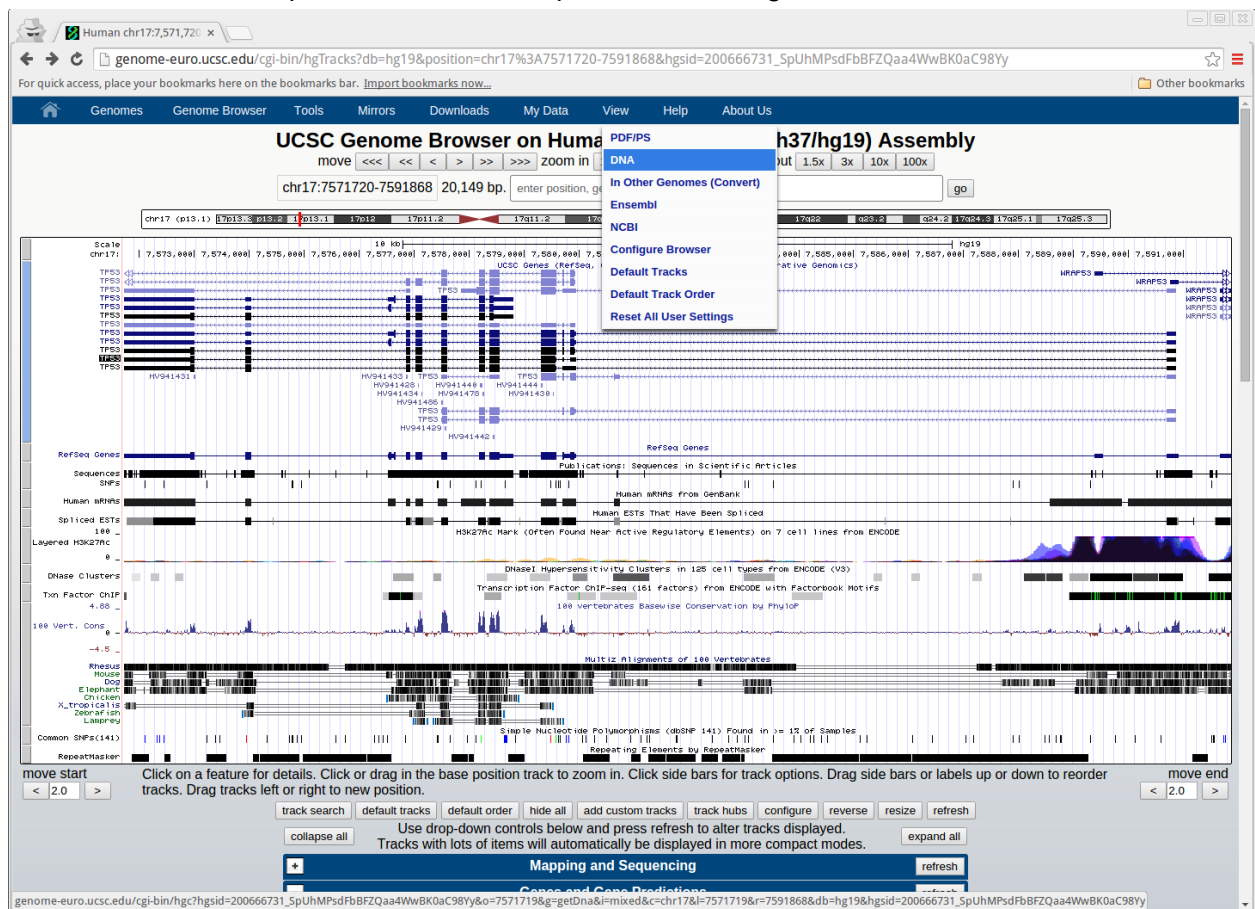
5 - Provare a cambiare l’ordine con cui le track vengono visualizzate nel “viewer”

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

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



## 8 - Visualizzare la Sequenza di DNA corrispondente alla regione nel "viewer".



Genomes Genome Browser Tools Mirrors Downloads My Data Help About Us

### Get DNA in Window (hg19/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.

#### 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)  
☒ get DNA. extended case/color options

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 UCSC Genes in rosso (254) e Spliced ESTs in verde (254). Cosa rappresentano le regioni colorate in giallo?

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

### Ricerche a partire da dati di sequenza

The screenshot shows the UCSC Genome Browser homepage. The 'Tools' menu is open, and the 'Blat' option is highlighted with a red circle. Below the menu, there is a section titled 'tools' with descriptions for various tools including Genome Browser, In-Silico PCR, LiftOver, and VisiGene. The background features a blue DNA double helix graphic.

1 - Usare BLAT per fare una ricerca di similarità a partire dalle seguenti 2 sequenze:

>Seq1

```
actggcgctaaaagttttgagctttcctttacaattctcaaaagtctagagccaccgtccaggagcaggtagct
gctgggctccggggacactttcggttcgggctgggagcgtgctttccacgacggtgacacgcttcctggattgg
cagccagactgccttcgggtcactgccatggaggagccgcagtcagatcctagcgtcgagccccctctgagtca
ggaaacattttcagacctatggaaactacttcctgaaaacaacgttctgtcccccttgccgtcccaagcaatgga
tgatttgatgctgtccccggacgatattgaacaatggttcactgaagaccaggtccagatgaagctcccagaat
```

gccagaggctgctcccccggtggcccctgcaccagcagctcctacaccggcggcccctgcaccagccccctcctg  
gccccctgtcatcttctgtcccttcccagaaaacctaccaggggacctacggtttccgtctgaggcttcttgattc  
tgggacagccaagtctgtgacttgcacgtactccccctgccctcaacaagatgttttgccaactggccaagacctg  
ccctgtgcagctgtgggttgattccacacccccgcccggctttacgcgccatggccatctacaagcagtcacagc  
acatgacggagggttgtagggcgctgccccaccatgagcgctgctcagatagcgatggctctggccccctcctcagc  
atcttatccgagtggaaaccttgatttgcgtgtggagtatttggatgacagaaacacttttcgacatagtgtggt  
ggtgcccttttaccatgctgaggttggctctgactgtaccacatccactcattaactacatgtgtaacagttcc  
tgcattgggaggcatgaaccggaggcccatcctcaccatcatcacactggaagactccagtggtaattctactggga  
cggaaacagctttgaggtgcgtgtttgtgcctgtcctgggagagaccggcgcacagaggaagagaatctccgaag  
aaaggggagcctcaccacgagctgccccaggaggactaagcgagcactgccaacaacaccag

>Seq2

cataccagcttagattttaagggttttactgtgagggatgtttgggagatgtaagaaatgttcttgagtttaagg  
gttagttttacaatcagccacattctaggtaggggcccacttcaccgtactaaccaggaagctgtccctcactgt  
tgaatcttcttaacttcaaggcccatactgtgaaatgctggcatttgcacctacctcacagagtgcattgggtt  
aatgaaataatgtacatctggccttgaaaccaccttttattacatggggcttagaactttaccatttaggattt  
agtttatcctctccctgaaattggctcgggtgggttggttagtcatctacagttgggcagctgggttaggttagaggag  
ttgtcaagtctctgctggcccagccaaacctgtctgacaacctcttggtgaaccttagtacctaaaaggaaatc  
tcaccccatcccacttcataccctggaggatttcatctcttgatatgatgatctggatccaccaagacttggtt  
tatgctcagggcaatttctgttttcttt  
gcccaggctggagtgagtggtgatcttggttactgctttaaacctttttaacttttaggggtttaccacctc  
agcctccggagtagctgggaccacacgggtcatgccaccatggccagccaacttttgcgttattaatgttttgta  
gagatggggctctcacagtgttgcccaggctggtctcaaacctcctgggctcaggcgatccacctgtctcagcctcc  
cagagtgtgggattacaattgtgagccaccacgtccagctggaagggtcaacatcttttacattctgcaagcac  
atctgcattttcacccacccttccccctccttctcccttttttatatcccatttttatatcgatctcttattttac  
aataaaactttgctgccacct

### Esercizio di riepilogo

- 1- Vogliamo ora identificare tutti gli SNP (dbSNP versione 142) Coding STOP GAINED nel gene NOTCH1 di topo e visualizzarli in rosso.
- 2- Visualizzare gli SNP (dbSNO versione 142) Coding Non-Synonymous, sempre nel gene NOTCH1 di topo e colorarli di rosso