

Bioinformatica II

LM Biologia Evoluzionistica, Università di Padova

Docenti: Dr. Giorgio Valle, Dr. Stefania Bortoluzzi

Esercitazione 1

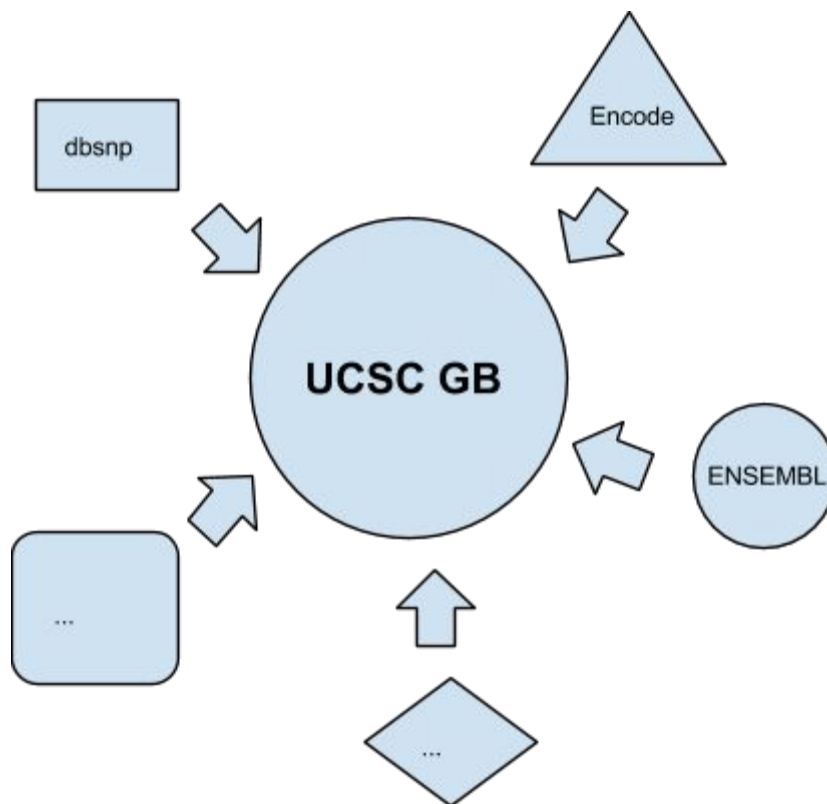
Padova, 10 novembre 2015

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.



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

Human (*Homo sapiens*) Genome Browser Gateway

The UCSC Genome Browser was created by the [Genome Bioinformatics Group of UC Santa Cruz](#).
Software Copyright (c) The Regents of the University of California. All rights reserved.

group	genome	assembly	position	search term
Mammal	Human	Feb. 2009 (GRCh37/hg19)	chr21:33,031,597-33,041,570	TP53


[Click here to reset](#) the browser user interface settings to their defaults.

[track search](#) [add custom tracks](#) [track hubs](#) [configure tracks and display](#)

Human Genome Browser – hg19 assembly (sequences)

The February 2009 human reference sequence (GRCh37) was produced by the [Genome Reference Consortium](#). For more information about this assembly, see [GRCh37](#) in the NCBI Assembly database.

Sample position queries



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

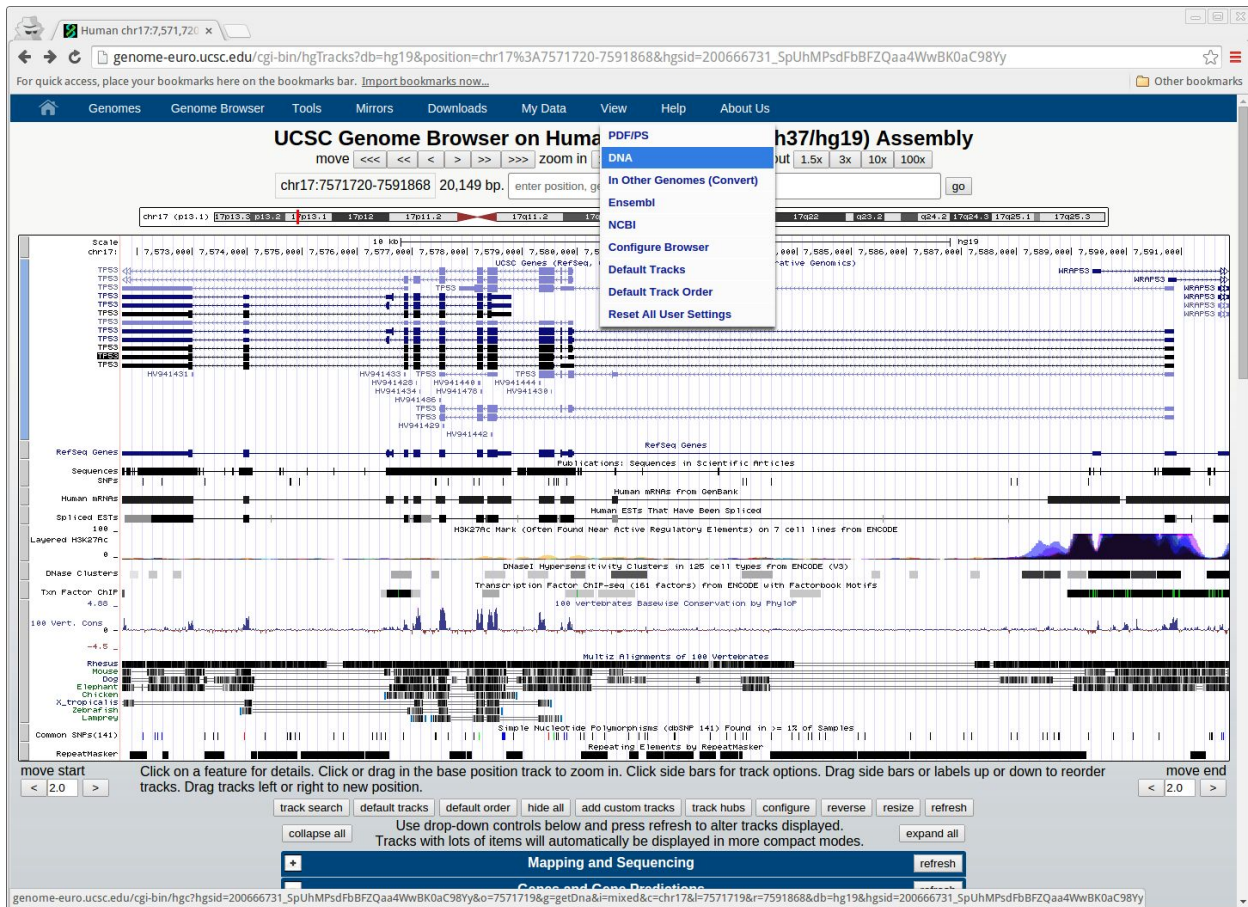
4 - Includere 1000 basi a valle (promotore putativo) del trascritto scelto: “chr17:7,571,720-7,591,868”

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.

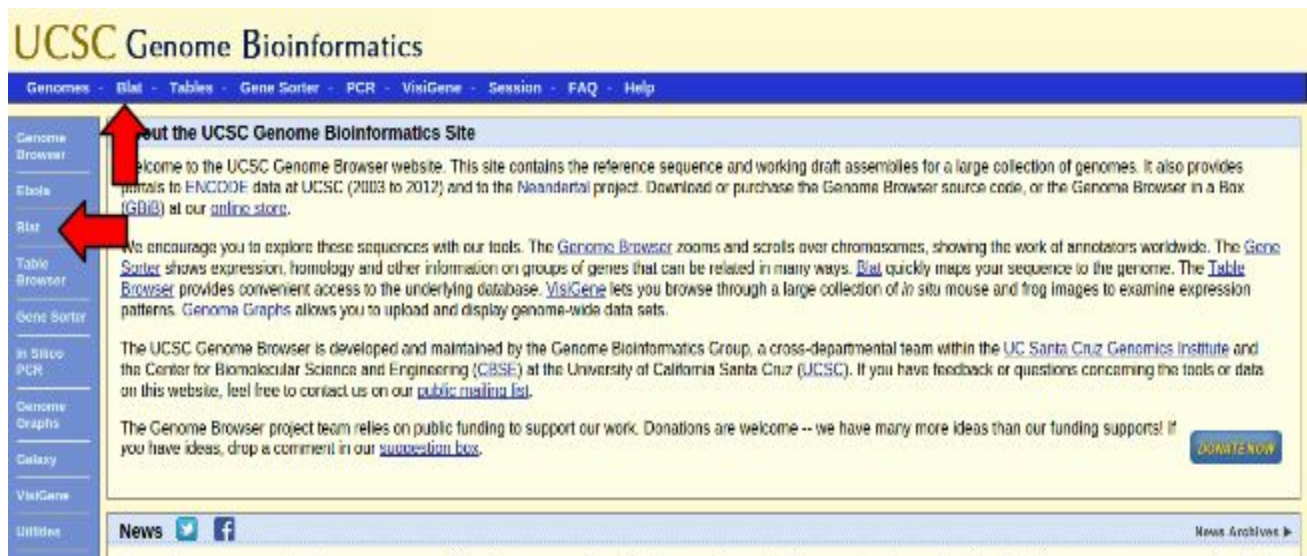
8 - Visualizzare la Sequenza di DNA corrispondente alla regione nel "viewer".



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



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

>Seq1

```
gatgggattgggggttttctttaaacgtgtgctcaagactggcgctaaaagttttgagcttctcaaaagtctagag
ccaccgtccaggaggagcaggtagctgctggggtccggggacactttgctgctcgggctgggagcgtgctttccacga
cgggtgacacgcttccctggattggcagccagactgccttccgggtcactgccatggaggagccgcagtcagatcc
tagcgtcgagccccctctgagtcaggaaacattttcagacctatggaaactacttctgaaaacaacgttctgtc
ccccttgccgtcccaagcaatggatgatttgatgctgtgtcccgacgatattgaacaatggttcactgaagacc
aggtccagatgaagctccagaatgccagaggctgctcccccggtggccctgcaccagcagctcctacaccggc
ggccctgcaccagccccctctggccctgtcatcttctgtcccttccagaaaacctaccaggggacctacgg
tttccgtctgggcttcttgattctgggacagccaagtctgtgacttgacgtactccctgccctcaacaagat
gttttgccaactggccaagacctgccctgtgcagctgtgggttgattccacaccccgccggcaccgcgctcg
cgccatggccatctacaagcagtcacagcacatgacggagggtgtgaggcgctgccccaccatgagcgctgctc
agatagcgatggtctggccccctcctcagcatcttatccgagtggaaaggaaatttgctgtgaggatattggatga
cagaaacattttcgacatagtgtggtggtgcccttttaccatgctgaggttggtctgactgtaccaccatcca
ctacaactacatgtgtaacagttcctgcatgggcggcatgaaccggaggccatcctcaccatcatcacactgga
agactccagtggtaatctactgggacggaacagctttgaggtgcgtgtttgtgcctgtcctgggagagaccggcg
cacagaggaagagaatctccgcaagaaaggggagcctcaccacgagctgccccaggaggagcactaagcgagcact
gccaacaacaccag
```

>Seq2

```
tttgttgtggggagggggatggggagtaggacataccagcttagattttaaggtttttactgtgagggatgtttg
ggagatgtaagaaatgttcttgagtttaagggttagtttacaatcagccacattctaggtaggggcccacttcac
cgtactaaccaggaagctgtccctcactgttgaattttctctaacttcaaggcccatatctgtgaaatgctggc
atttgcacctacctcacagagtgcattgtgagggttaatgaaataatgtacatctggccttgaaaccacctttta
```

ttacatggggtctagaactttaccatttaggatttagtttatcctctccctggtgggtcggtgggttggtagttt
ctacagttgggcagctggttaggtagagggagttgtcaagtctctgctggcccagccaaaccctgtctgacaacc
tcttggtgaaccttagtacctaaaaggaaatctcaccccatccacaccctggaggatttcattctcttgatatg
atgatctggatccaccaagacttgttttatgctcagggtcaatttctttttctttttttttttttttttttttt
ttctttgagactgggtctcgctttgttgcccaggctggagtgagtggtgatcttggttactgctttaaac
tttttaacttttagggtttaccacctcagcctccggagtagctgggaccacaggttcatgccaccatggccagcc
aacttttgcatgtttttagagatggggtctcacagtgttgcccaggctggtctcaaactcctgggctcaggcga
tccacctgtctcagcctccagagtgttgaggattacaattgtgagccaccacgtccagctggaagggtcaacatc
ttttacattctgcaagcacatctgcattttcacccaccctcccctccttctccctttttatatcccattttta
tatcgatctcttattttacaataaaactttgctgccacctgtgtgtctgaggggtg

Esercizio di riepilogo

Vogliamo ora identificare tutti gli SNP Coding - Non-Synonymous nel gene NOTCH1 di topo e visualizzarli in rosso.

Common SNPs(138) Track Settings

Simple Nucleotide Polymorphisms (dbSNP 138) Found in >= 1% of Samples ([All Variation and Repeats tracks](#))

Display mode: dense

Include observed alleles in name: ☐

Show alleles on strand of reference genome reported by dbSNP: ☐

☒ Use Gene Tracks for Functional Annotation

☒ Filtering Options

☒ Coloring Options

SNP Feature for Color Specification: Function

The selected "Feature for Color Specification" above has the selection of colors below for each attribute. Only the color options for the feature selected above will be used to color items; color options for other features will not be shown. If a SNP has more than one of these attributes, the stronger color will override the weaker color. The order of colors, from strongest to weakest, is red, green, blue, gray, and black.

Unknown	black	Locus	black	Coding - Synonymous	black	Coding - Non-Synonymous	red
Untranslated	black	Intron	black	Splice Site	black		

[View table schema](#)

Data last updated: 2014-01-17