

# Bioinformatica II

LM Biologia Evoluzionistica, Università di Padova

Docente: Prof. Stefania Bortoluzzi

## Esercitazione 1

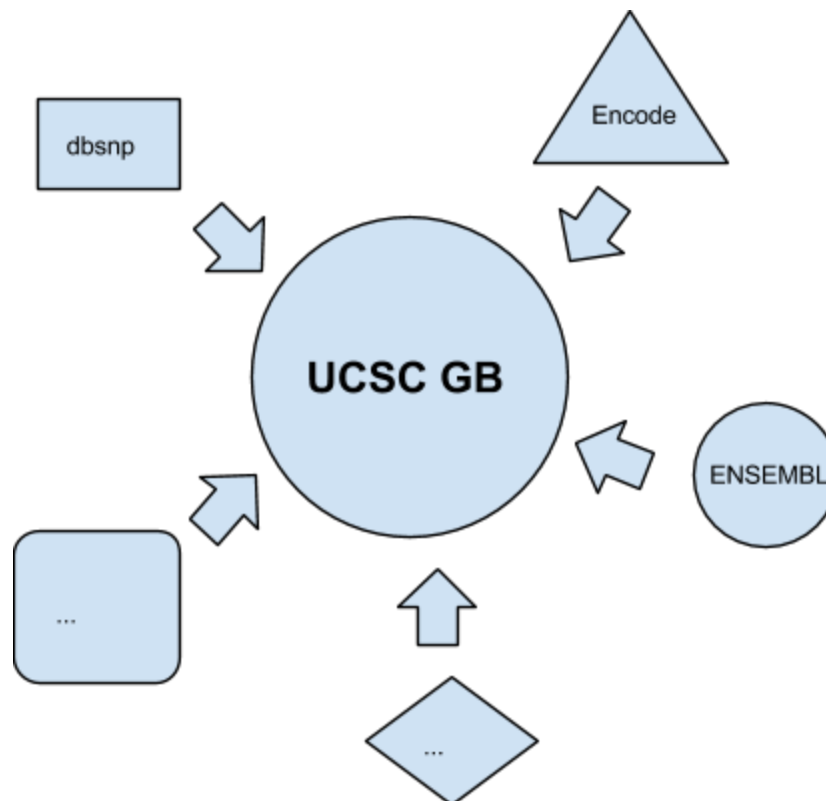
Padova, 2 novembre 2017

### GUIDA

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

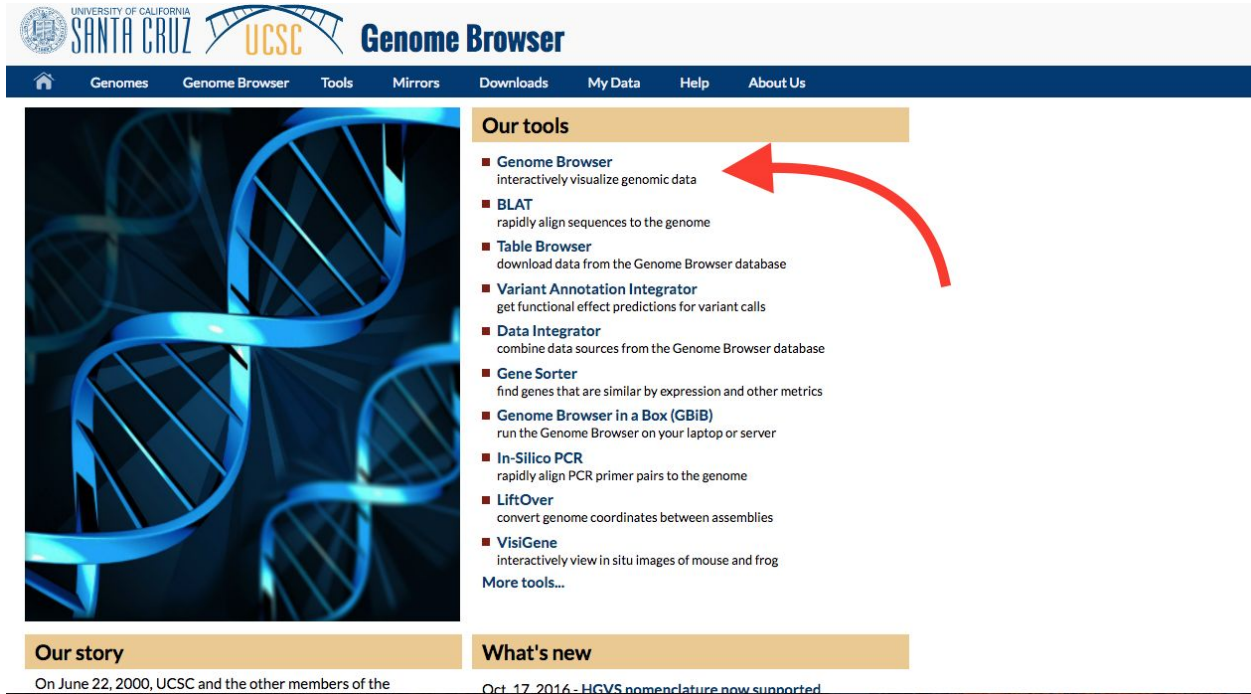
##### Obbiettivo dell'esercitazione

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.



UCSC Genome Browser

Genomes Genome Browser Tools Mirrors Downloads My Data Help About Us

**Our tools**

- **Genome Browser**  
interactively visualize genomic data
- **BLAT**  
rapidly 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...](#)

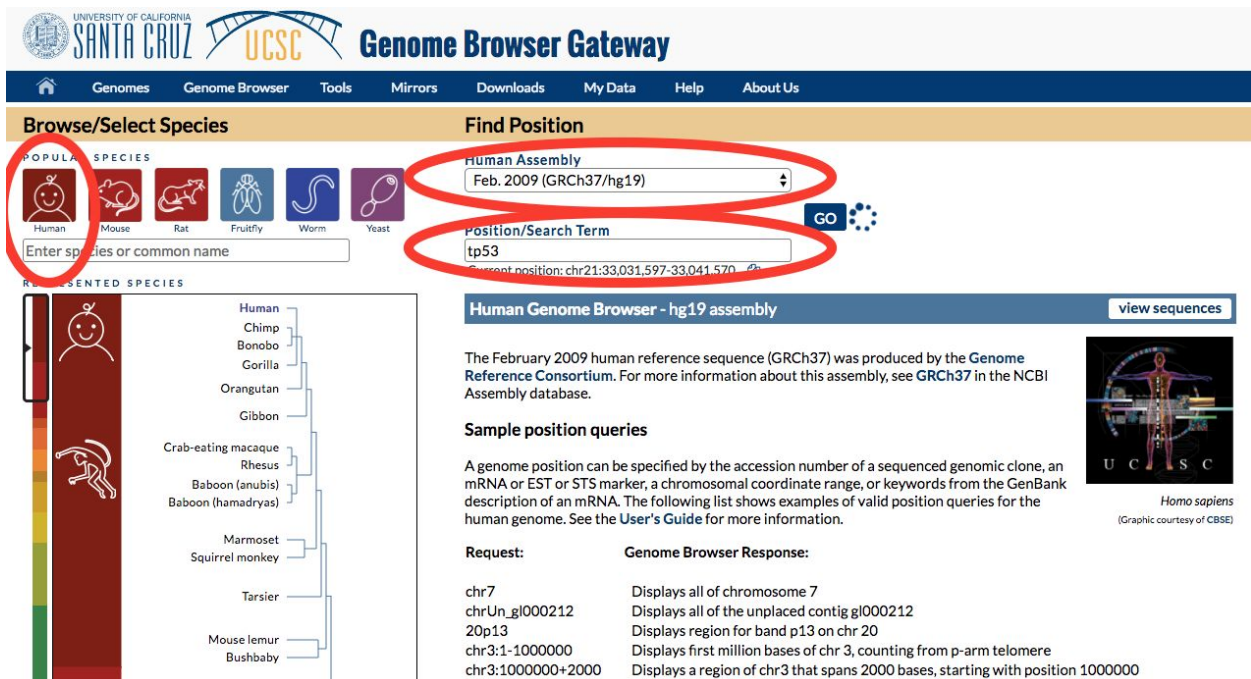
**Our story**

On June 22, 2000, UCSC and the other members of the

**What's new**

Oct. 17, 2016 - HGVS nomenclature now supported

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



UCSC Genome Browser Gateway

Genomes Genome Browser Tools Mirrors Downloads My Data Help About Us

**Browse/Select Species**

POPULAR SPECIES

Human Mouse Rat Fruitfly Worm Yeast

Enter species or common name

**Find Position**

Human Assembly  
Feb. 2009 (GRCh37/hg19)

Position/Search Term  
tp53

GO

**Human Genome Browser - hg19 assembly**

view 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**

A genome position can be specified by the accession number of a sequenced genomic clone, an mRNA or EST or STS marker, a chromosomal coordinate range, or keywords from the GenBank description of an mRNA. The following list shows examples of valid position queries for the human genome. See the User's Guide for more information.

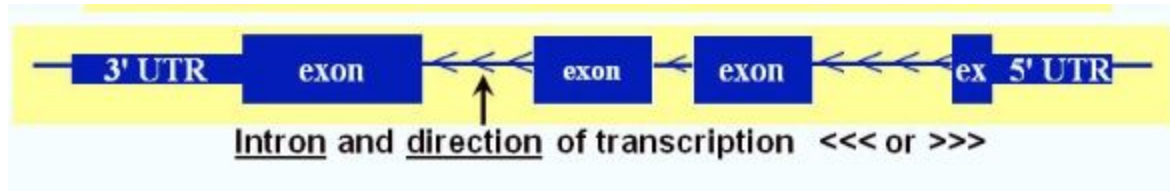
Request:	Genome Browser Response:
chr7	Displays all of chromosome 7
chrUn_gi000212	Displays all of the unplaced contig gi000212
20p13	Displays region for band p13 on chr 20
chr3:1-1000000	Displays first million bases of chr 3, counting from p-arm telomere
chr3:1000000+2000	Displays a region of chr3 that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

Represented Species

Human  
Chimp  
Bonobo  
Gorilla  
Orangutan  
Gibbon  
Crab-eating macaque  
Rhesus  
Baboon (anubis)  
Baboon (hamadryas)  
Marmoset  
Squirrel monkey  
Tarsier  
Mouse lemur  
Bushbaby

Homo sapiens  
(Graphic courtesy of CBSE)

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

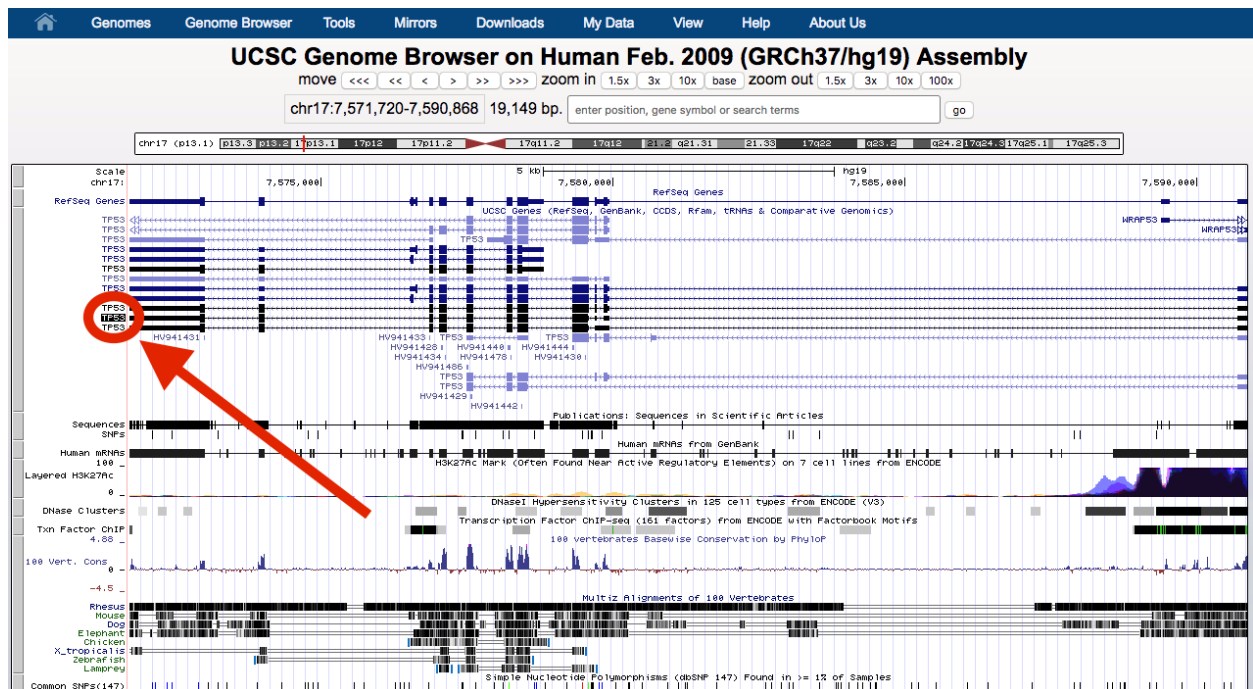


4 - Includere 1000 basi a valle (promotore putativo) del trascritto scelto: “chr17:7,571,720-7,591,868”

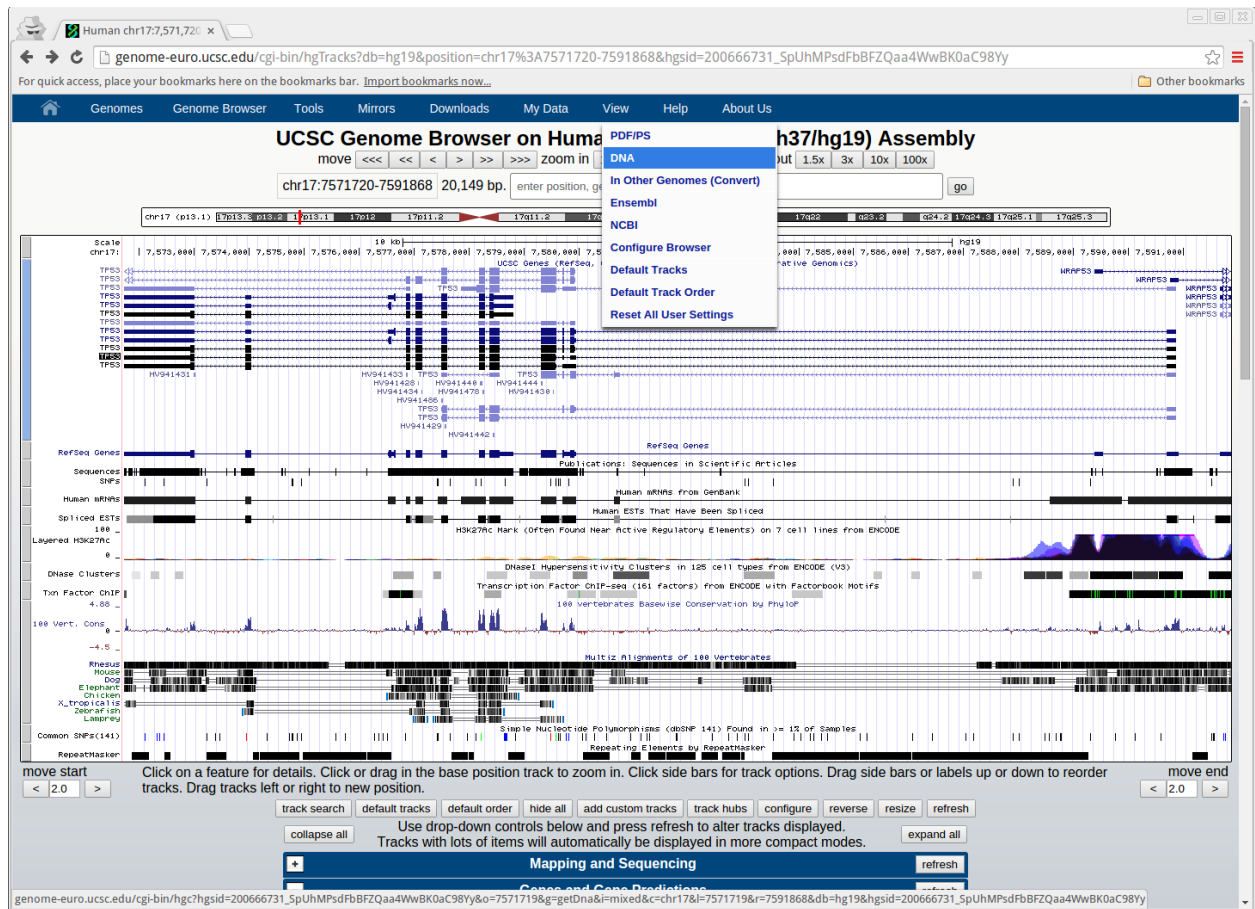
5 - Provare a cambiare l’ordine con cui le track vengono visualizzate nel “viewer”

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.

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, displaying various tools for genomic analysis. The 'Blat' tool is circled in red, and the 'Data Integrator' tool is also circled in red. The 'Genome Browser' tool is highlighted in orange. The 'What's new' section shows updates from October 17, 2016, and August 31, 2016.

**Tools**

- Blat
- Table Browser
- Variant Annotation Integrator
- Data Integrator
- Gene Sorter
- Genome Graphs
- In-Silico PCR
- LiftOver
- VisiGene
- Other Utilities

**Genome Browser**

- Interactively visualize genomic data
- Align sequences to the genome
- Download data from the Genome Browser database
- Variant Annotation Integrator
- Functional effect predictions for variant calls
- Data Integrator
- 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.

**What's new**

Oct. 17, 2016 - **HGVs nomenclature now supported**

Aug. 31, 2016 - **Genome Browser for Chicken**

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

>Seq1

```
actggcgctaaaagttttgagctttcctttacaattctcaaaagtctagagccaccgtccagggagcaggtagctgc
tgggctccggggacactttgcgttcgggctgggagcgtgctttccacgacggtgacacgcttccctggattggcagc
cagactgccttcgggtcactgccatggaggagccgcagtcagatcctagcgtcgagccccctctgagtcaggaaac
attttcagacctatggaaactacttctgaaaacaacgttctgtcccccttgccgtccaagcaatggatgatttga
tgctgtccccggacgatattgaacaatgggttcactgaagaccaggtccagatgaagctcccagaatgccagaggct
```

gctcccccggtggcccctgcaccagcagctcctacaccggcggcccctgcaccagccccctcctggcccctgtcatc  
ttctgtcccttcccagaaaacctaccaggggacctacgggtttccgtctgggcttcttgcattctgggacagccaagt  
ctgtgacttgcacgtactccccctgccctcaacaagatgttttgccaactggccaagacctgccctgtgcagctgtgg  
gttgattccacacccccgccggctttacgcgccatggccatctacaagcagtcacagcacatgacggagggttgta  
ggcgctgccccaccatgagcgctgctcagatagcgatggcttgcccctcctcagcatcttatccgagtggaaacc  
ttgatttgcgtgtggagtatttggatgacagaaacacttttcgacatagtgtggtggtgcccttttaccatgctgag  
gttggctctgactgtaccacatccactcattaactacatgtgtaacagttcctgcatggggcgcatgaaccggagg  
cccatcctcaccatcatcacactggaagactccagtggtaatctactgggacggaacagctttgagggtgcgtgtttg  
tgcctgtcctgggagagaccggcgacagaggaagagaatctccgcaagaaaggggagcctcaccacgagctgcccc  
cagggagcactaagcgagcactgcccacaacaccag

>Seq2

cataccagcttagattttaagggtttttactgtgagggatgtttgggagatgtaagaaatgttcttgcagttaagggt  
tagtttacaatcagccacattctaggtagggggccacttcaccgtactaaccagggaagctgtccctcactgttgaa  
ttttctctaacttcaaggccatatctgtgaaatgctggcatttgcacctacctcacagagtgcattgggttaatgaa  
ataatgtacatctggccttgaaaccaccttttattacatggggctctagaactttaccatttaggatttagtttattc  
ctctccctgaaattggctgggtgggttggttagtcatctacagttgggcagctgggttaggttagggagttgtcaagtc  
tctgctggcccagccaaaccctgtctgacaacctcttggtgaaccttagtacctaaaaggaaatctcaccccatccc  
acttcataccctggaggatttcatctcttgtatatgatgatctggatccaccaagacttgttttatgctcagggtca  
atttctgttttcttt  
agtggcgtgatcttggcttactgctttaaacctttttaacttttaggggtttaccacctcagcctccggagtagctgg  
gaccacacggttcatgccaccatggccagccaacttttgcgttattaatgttttgtagagatggggctctcacagtgt  
tgcccaggctgggtctcaaactcctgggctcaggcgatccacctgtctcagcctccagagtgtgggattacaattg  
tgagccaccacgtccagctggaagggtcaacatcttttacattctgcaagcacatctgcattttcacccacccttc  
ccctccttctcccttttttatatcccattttttatatcgatctcttattttacaataaaaactttgctgccacct

### 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