

Corso di Bioinformatica

http://compgen.bio.unipd.it/~stefania/Didattica/AA2010-2011/Bioinformatica_BTS/Bioinfo_BTS.html

Per ogni punto dell'esercitazione copiare ed incollare le informazioni richieste in un file sotto forma di breve testo con figure, in modo da assemblare un traccia di relazione con brevi commenti.
Alla fine inviare via mail la relazione prodotta nel corso dell'esercitazione (stefania.bortoluzzi@unipd.it).
Salvare il file della relazione ogni 5'!

IV ESERCITAZIONE

Localizzazione e ricostruzione della struttura di un gene a partire da un cDNA (BLAT, Genome Browser, EntrezGene e OMIM)

La sequenza completa del genoma umano è disponibile ed annotata, e' estremamente importante imparare a "navigare" questo genere di informazioni e poterle sfruttare per gli scopi della propria ricerca. In questa esercitazione cercheremo di identificare e localizzare un gene a partire solo dalla sequenza di un breve cDNA ad esso afferente, di ricostruirne la struttura mediante BLAT e Genome Browser. Recupereremo poi alcune informazioni importanti riguardo al gene stesso, al suo prodotto proteico e alle malattie genetiche correlate utilizzando EntrezGene e OMIM.

1 Collegarsi all'home page <http://genome.ucsc.edu/> ed al programma BLAT.

2 Scegliere l'ultima versione del genoma umano.

3 Incollare la sequenza seguente (cut and paste):

```
>(EST) Human retina cDNA Tsp509I-cleaved sublibrary Homo sapiens cDNA not directional, mRNA sequence
TNNNNNNGCCNACCAAGGTGAATTGCCNTGTGGGCCTTGGTGGTCTGGACATCGAGCGGTACGTGGTG
GTGTGTAAGCCCATGAGCAACTTCCGCTTCGGGGAGAACCATGCCATCATGGCGTTGGCTTACCTGGG
TCATGGAGCTGGGCTGCGNCGAACCCCACTCGCCGGTGGTCCAGGTACATCCCCGAGGGCTGCAGTG
CTCGTGTGGAATCGACTACTACACGNTCAAGCCGGAGGTCAACAACGAGTCTTTTGTTCATCTACATGTTT
GNGGTCCACTTANCATCCCCATGATTATCATCTTTTTCTGCTATGGGCAGCTCGTCTTCANCGTCANGG
AGGCCGCTGNNCAGCAGCAGGAGTCAGCCACCACACAGAAGGCAGAGAAGGAGGTACCCCGCATGGTCAT
CATCATGGCATCGCTTCTGTGTTGGTGCCTACGCAGGGTGGATCTCATTTACCACAGGGTCAATT
CGGTCCATTTTCANGACATCAGGGTCTTGGCAAGAGGGCGCACTTNACTGNANTTTTATATGACAAGAG
NCCGATNATTTAACACACTTTGGGANACNCTGGGAGANGGGCNTTCCCTCAAAGGGAGACAGNCCCGCTA
NCCNCTGGTTTNGCCTTTGGTCTTCTNACCCCNNGNAAANCCAGGAGGCTTNAAGTCCAGAAATGTCNAC
NTNGGTTCCGNAANAAAAA
```

4 Inviare la ricerca.

5 Analizzare il risultato della ricerca di similarita'.

- **Quale regione del genoma risulta simile alla sequenza in analisi ?**
- **Con che percentuale di identita'?**

6 Cliccare su "browser" e considerare la situazione, dopo aver impostato la visualizzazione in modalita' "full" di bande cromosomiche, geni noti, RefSeq, human mRNA e EST, CpG islands, conservation.

- **Di quale gene fa parte la sequenza EST in analisi?**
- **Di quanti esoni si compone il gene e qual'e' la sua estensione nel genoma (regione trascritta)?**
- **Quanto e' lungo il primo esone?**

7 Predire la posizione del TSS piu' probabile e quindi "inquadrare" la regione di DNA genomico compresa tra le 1000 bp a monte del TSS e la fine dell'ultimo esone.

8 Recuperare e **salvare in un file di testo la sequenza di DNA corrispondente alla regione inquadrata** (in totale dovrebbe comprendere 1000 bp piu' la lunghezza del primo esone) utilizzando "Extended DNA Case/Color Options" per tenere traccia degli esoni e degli introni.

9 Inquadrare solo la regione del promotore comprendente le 1000 basi upstream al TSS. Quanti SNPs (SNPs 130) sono noti in questa regione?

10 A partire dalla pagina di descrizione "RefSeq Gene" della Rhodopsina, seguire il collegamento a Entrez Gene. Analizzare l'entry in dettaglio e cercare di rispondere alle seguenti domande:

- **Quali malattie genetiche sono causate da mutazioni del gene per la Rhodopsina?**
- **Qual e' il ruolo funzionale della proteina?**

11 A partire dalla pagina di descrizione "RefSeq Gene" della Rhodopsina, seguire il collegamento a OMIM. Considerare l'entry 180380.

- **In base al campo "allelic variants", che tipo di mutazione e' quella descritta dalla variante 180380.0003 ?**
- **In base alla "Clinical Synopsis", come si ereditano le patologie legate a mutazioni del gene per la Rhodopsina e quando si sviluppa la malattia?**