

### **Corso di Bioinformatica**

[http://compgen.bio.unipd.it/~stefania/Didattica/AA2011-2012/Bioinformatica\\_BTS/Bioinfo\\_BTS.html](http://compgen.bio.unipd.it/~stefania/Didattica/AA2011-2012/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 commenti.  
Alla fine inviare via mail la relazione prodotta nel corso dell'esercitazione ([stefania.bortoluzzi@unipd.it](mailto:stefania.bortoluzzi@unipd.it)).  
Salvare il file della relazione ogni 5'!

## **I ESERCITAZIONE**

### **Parte A: Data retrieval da ENTREZ**

ENTREZ (<http://www.ncbi.nlm.nih.gov/Entrez/index.html>) è un'importante sistema per la ricerca di informazioni biologiche che fornisce un'interfaccia unificata per accedere a molti database differenti per contenuto ed organizzazione.

#### **1 Ricerca bibliografica in ENTREZ Books:**

- Dare un'occhiata a "Book List".
- Cercare Cadherins in Cancer medicine.

#### **2 Ricerca bibliografica in ENTREZ PubMed:**

- Trovare l'articolo più recente sul rhabdomyosarcoma embrionale (embryonal rhabdomyosarcoma[Title]).
- Trovare l'articolo più recente su Science riguardante la produzione microRNA e angiogenesi (angiogenesis).

#### **3 Retrieval di sequenze con ENTREZ Nucleotide:**

- Cercare una sequenza del gene mitocondriale per il citocromo b di qualsivoglia specie di storione (acipenser, cytb).

#### **4 Retrieval di sequenze con ENTREZ Protein:**

- Cercare la sequenza della PROTEINA umana nebulin, annotarsi la lunghezza e copiare la sequenza in formato FASTA nel file Word.

#### **5 Entrez Genome:**

Consultare le seguenti pagine, a da ciascuna provare a raggiungere le informazioni relative ad un diverso genoma completo:

- Entrez genome project (<http://www.ncbi.nlm.nih.gov/genome/>) → *Propionibacterium acnes* KPA171202. Quanti geni protein coding contiene questo genoma batterico?
- Map Viewer (<http://www.ncbi.nlm.nih.gov/projects/mapview/>) → invertebrates → insects → *D. melanogaster* seguire i diversi link fino ad "entrare" mediante Mapview nel cromosoma X, osservare le diverse mappe mostrate in parallelo. Quanti geni sono contenuti nel cromosoma? Dalla tabella dei protein coding genes dedurre le coordinate del gene per la "*gamma-tubulin ring protein 84*" e scaricare la sequenza della proteina codificata dal trascritto A, in formato FASTA.
- Genomes (<http://www.ncbi.nlm.nih.gov/Genomes/>) → Human Genome Resources (<http://www.ncbi.nlm.nih.gov/projects/genome/guide/human/>). Usando Map Viewer, scaricare la sequenza della regione "D loop" del genoma mitocondriale umano in formato FASTA, a partire dal formato GenBank della sequenza stessa.

## **Parte B: Ricerca in Omim, Entrez Genes, GeneCards, Human Gene Mutation Database.**

Prenderemo in considerazione una malattia genetica e cercheremo di utilizzare i dati contenuti in diversi database (di malattie, geni, proteine e genomi) per reperire informazioni sul fenotipo, sui geni, sulle mutazioni che lo causano, ecc.

- 1) Collegarsi a OMIM e cercare “RETINITIS PIGMENTOSA”
  - Analizzare i risultati e commentarli (entries di geni; entries di malattie; ci sono piu' malattie diverse ?).
- 2) Scegliere “RETINITIS PIGMENTOSA 9”
  - Analizzare le informazioni sulla patologia (pagina principale, clinical synopsis e gene map) per capire di che malattia si tratta, come si eredita e dove mappa il locus associato.
- 3) Seguire il link “RP9 gene (607331)”
  - Di che gene si tratta ?
  - Ci sono delle varianti alleliche con probabile ruolo patogenetico? Sono legate al DBSNPs? E' possibile sapere quanti SNPs missenso sono stati descritti nel primo esone di questo gene, utilizzando l'informazione contenuta in DBSNP? (usare GeneView, selezionando “in gene region”)
- 4) Seguire il link a “EntrezGene” e guardare la pagina relativa al gene;
  - quanti esoni ci sono ?
  - quali sono i due geni “vicini” sul DNA genomico ?
  - a partire dal campo “NCBI Reference Sequences (RefSeq)”, scaricare la sequenza della proteina codificata in formato FASTA (NP\_976033.1).
- 5) Collegarsi a GeneCards e cercare “RP9”:
  - Analizzare la pagina di descrizione del gene, con particolare attenzione ai seguenti campi:
  - “Proteins” (seguire il link ad Uniprot e poi a “Family and domain databases, SMART Graphical view of domain structure”);
  - “Expression in Human Tissues”;