

I Esercitazione di Bioinformatica
“Data retrieval”

Seguire il percorso indicato nel testo sottostante e stilare una breve relazione con i contenuti recuperati e man mano elaborati usando diverse risorse informatiche.

A) Data retrieval da database dell'NCBI

NCBI (<http://www.ncbi.nlm.nih.gov/>) è un'importante risorsa da cui reperire informazioni biologiche. Fornisce un'interfaccia unificata per accedere a molti database differenti per contenuto e organizzazione, quali database di conoscenza (letteratura, geni, malattie), biosequenze (nucleotidiche, proteiche) e strutture (3D di proteine, domini proteici conservati).

A partire da <http://www.ncbi.nlm.nih.gov/guide/all/>

1) Ricerca bibliografica in NCBI Bookshelf (Resources, Literature):

- Dare un'occhiata ai libri presenti usando con i rispettivi filtri (*ad es. All subjects, Books, poi Neuroscience*).
- Cercare Cadherins nel libro Cancer medicine e identificare un capitolo di interesse da cui trarre un breve testo.

2) Ricerca bibliografica in NCBI PubMed (Literature):

- Usare PubMed Advanced Search Builder Imparare a usare:
 1. la ricerca mirata in specifici campi delle entry (ad esempio controllare la differenza tra cercare solo “pox” e “pox[Title]”);
 2. e a combinare le ricerche con gli operatori Booleani (ad esempio controllare la differenza tra “pox[Title]” e “pox[Title] NOT chicken[Title]”);
 3. a usare i filtri per data, tipo di pubblicazione ecc.
- Trovare l'articolo più recente sul rhabdomyosarcoma embrionale (embryonal rhabdomyosarcoma [Title]).
- Trovare l'articolo più recente che abbia tra gli autori “Rajewsky-N”.
- Trovare un articolo pubblicato nel 2012 sulla rivista Nature medicine riguardante l'autismo (autism)

3) Retrieval di sequenze con NCBI Nucleotide (DNA & RNA; Genome, gene and transcript sequence data):

- Cercare la sequenza di riferimento (Refseq) del gene mitocondriale per il citocromo b umano (cytochrome; cytb)
- Quante sequenze ci sono?
- Provare a partire dal database “gene”
- Scaricare la sequenza del gene in formato FASTA. Quanto è lunga?

4) Retrieval di sequenze con NCBI Protein:

- Cercare la sequenza di riferimento della PROTEINA umana alpha crystallin-A, annotarsi la lunghezza, copiare la sequenza in formato FASTA e considerare la visualizzazione “Graphics” dell’entry.

5) NCBI Genome:

Consultare le pagine raggiungibili da NCBI Genome (<http://www.ncbi.nlm.nih.gov/genome>), quali “Human genome”, “Viruses”, ecc.

Provare a raggiungere le informazioni relative ad uno specifico genoma completo:

→ *Propionibacterium acnes* KPA171202

Usare Taxonomy database per la classificazione (classe, sottoclasse, ecc.) utile alla navigazione dell’albero.

Quanti geni protein coding contiene questo genoma batterico?

Scaricare la seq in format FASTA della proteina più lunga codificata.

→ *Homo sapiens* (human) genome view, Annotation Release 107. Raggiungere le mappe del cromosoma Y (usare ideogrammi cromosomi), e cercare di capire quanto è lungo e quanti geni annotati contiene. Usare “Table view” per identificarne alcuni inclusi nella regione di 4 megabasi adiacente a YpTer (agire su region shown) e circa corrispondente alla banda Yp11.3 e scaricare le descrizioni dei geni mediante “Download data”.

Dalla tabella seguire il link a “LINC00278”, orientarsi e scaricare la sequenza di riferimento (Accession NR_...) in format FASTA.

B) Dalla malattia genetica, al gene, attraverso il genoma.

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”
- 2) Scegliere “RETINITIS PIGMENTOSA 9”
 - a. Analizzare velocemente le informazioni sulla patologia (pagina principale, clinical synopsis e gene map) per capire di che malattia si tratta, come si eredita e in quale cromosoma e regione genomica si trova il locus associato.
- 3) Seguire il link al locus e quindi alla “location” per arrivare alla visualizzazione grafica nel **UCSC Genome Browser** del contenuto di informazione della regione del genoma umano che contiene il gene.
 - Considerare la track del gene (colore blu). Di che gene si tratta? Di quanti esoni è composto il gene?
 - Quali sono i due geni “vicini” sul DNA genomico ?
 - Cliccando sulla track del gene considerare la pagina relative. A partire dal campo “Sequence and Links to Tools and Databases” scaricare la sequenza della proteina codificata.