ESERCITAZIONE 1

OBIETTIVO: L'esercitazione prevede l'utilizzo di risorse web per la ricerca di informazioni di carattere biologico relative al **gene**della β -globina.

La ricerca dovrà essere effettuata a partire dalle seguenti banche dati integrate:

http://www.ncbi.nlm.nih.gov/Entrez	NCBI - Entrez
http://srs.ebi.ac.uk	EBI -SRS6

e dalla banca dati genomica: http://www.ensembl.org

ENSEMBL

Entrare in ciascuno di questi siti e salvarli nei preferiti (dal menu' Preferiti/Favoriti selezionare aggiungi a preferiti/add to favorites).

1) Sito dell'NCBI <u>http://www.ncbi.nlm.nih.gov/Entrez</u>

Cliccare su **Nucleotide** e, nella pagina che si aprira', digitare nella finestra di testo: **globin AND beta** (NB: AND operatore booleano, va in maiuscolo) e cliccare su GO. Si ha un numero eccessivo di risultati.

Per restringere il numero di entries del database individuati cliccare su **Preview/Index** sotto la casella di testo, vi sara' chiesto di sciegliere tra tre tipi di raggruppamenti di databases di sequenze nucleotidiche: Expressed Sequence Tag (EST), sequenze genomiche estese (GSS, Genome Survey Sequence) e CoreNucleotide, cioe' tutto cio' che rimane. Scegliere **CoreNucleotide** cliccando sulla voce corrispondente. Quindi digitare nuovamente **globin AND beta** nella finestra di testo (questa informazione potrebbe essere stata persa) e

selezionare dal menu in fondo alla pagina **Organism**, inserire nella casella a fianco la parola chiave **homo** e successivamente cliccare sul tasto **AND** (sotto la casella dove è stato scritto homo). A questo punto la query così composta(**beta AND globin and homo[Organism]**) apparirà sulla casella di testo superiore (controllate!) e si dovrà cliccare ancora su **GO** per inviare la richiesta al database. In questo modo la query viene ristretta all'*homo sapiens*. (vedi **Figura** che segue).



Al fine di restringere ulteriormente il numero dei entries trovati cliccare su Limits posto sotto la casella di testo superiore (a fianco di Preview/Index); a questo punto selezionare il checkbox "exclude all of the above" all'estrema destra e dal menu a tendina, denominato "Molecule", selezionare "mRNA". Significa che stiamo selezionando *solo* l'RNA messaggero ed escludendo banche dati poco utili ai nostri fini, come ad esempio quelle per STS (sequence tagged sites, pezzi di sequenze genomiche che servono a costruire le mappe fisiche dei cromosomi). Cliccare su GO ed ottenere i risultati.

I risultati ottenuti (quanti?) sono in numero accettabile ed è possibile analizzarli manualmente alla ricerca della Beta globina a cui siamo interessati.

Una delle entries (o voci) che contiene piu' informazioni sulla Beta globina, in altre parole che e' annotata con maggiore cura, corrisponde all'ID **NM_000518** (seconda pagina di risultati). Cliccando su di esso sara' possibile accedere ad una serie di informazioni utili sul gene relativo.

Tra le molte informazioni disponibili e da esplorare sono presenti:

- voci di Pubmed (referenze bibliografiche) da consultare, riguardano informazioni dettagliate e sperimentali sul gene e sul corrispondente prodotto genico;

- voci di OMIM (Online Mendelian Inheritance in Man) sulle malattie genetiche umane associate alla Beta globina: (cliccare sui link che iniziano con MIM:);

- voci di GeneID per informazioni sul gene e il cromosoma: (cliccare sui link GeneID). Appuntate la localizzazione del gene nel genoma umano.

-- voci di protein che da' accesso alla entry per la corrispondente sequenza proteica (cliccare su protein_id).

Ora che la entry (o voce) di nostro interesse e' stata trovata grazie all'uso di giusti indici (o keyword), e conosciamo quindi il suo codice identificativo, possiamo avere un quadro di insieme di tutto cio' che e' associato da Entrez a questa entry nelle diverse banche dati primarie. Semplicemente caricate nuovamente la pagina di Entrez e digitate il codice **NM_000518** nella casella di testo, quindi cliccate su **GO** (come nella **Figura** che segue). Notate se avete perso il collegamento a qualche informazione interessante precedentemente trovata.



2) Sito dell'EBI - SRS6 http://srs.ebi.ac.uk

Cliccare in alto a sinistra su Library Page. Una volta entrate/i in questa pagina, selezionare EMBL tra le Nucleotide sequence databases. Quindi cliccare sul bottone giallo a sinistra:

Extended Query Form (modulo per ricerca estesa).

Inserire nelle caselle di testo le parole chiave per la ricerca:

nel campo "**AllText**" digitare "**beta & globin**" (notare la differenza nell'operatore booleano rispetto all'NCBI & al posto di AND).

Quindi selezionare "**mrna**" in corrispondenza di "**molecule**" e "**hum**" in corrispondenza del menu "**Division**" (parte del database contenente solo sequenze umane escluse le ESTs).

Impostata la query inviarla cliccando sul tasto "**search**" in alto a destra. Si otterranno poche voci (quante?) tra cui EMBL:HSBGL1, che contiene le informazioni desiderate.

Confrontare le informazioni ottenute dalla query all'NCBI e quella all'EBI.

3) Ensembl http://www.ensembl.org

Entrare nel sito di Ensembl. Prendere nota di quanti genomi sono stati sequenziati e resi disponibili alla comunita' scientifica ad oggi. Quindi, selezionare il genoma umano cliccando su "Homo sapiens", selezionare poi "gene" dal menu' a tendina e digitare "globin AND beta", quindi cliccare su GO (vedi Figura che segue).



Cliccando sulla prima entry si puo' accedere a tutte le informazioni disponibili relative al gene della globina beta nel contesto del genoma umano.

Il sito e' molto ricco e complesso, cerchiamo di cogliere le informazioni piu' rilevanti.

Scorrendo la pagina nella box **Orthologue prediction** troviamo i link ad un elenco di geni ortologhi, ovvero geni che svolgono un'analoga funzione in organismi differenti. Cliccando su **align** e' anche possibile visualizzare l'allineamento tra le sequenze proteiche nell'uomo e, ad es., nel topo e accedere alle entry relative ai geni ortologhi trovati, nei relativi cromosomi. Scegliete una specie come esempio e annotate la similarita' delle sequenze proteiche e la diversa localizzazione cromosomica.

A seguire, nella box **Disease Matches** troviamo link alla banca dati OMIM, con informazioni circa le mutazioni patogene del gene.

Nella box **GO** troviamo link alla banca dati **G**ene **O**ntology, che contiene informazioni sulla funzione del prodotto genico.

Nella box **Protein family** troviamo un link ad altre entries relative a geni umani omologhi (paraloghi) presenti sullo stesso cromosoma.

Nella box **Transcript structure** troviamo una rappresentazione dell'organizzazione del gene in introni ed esoni

Nella box **Protein feature** troviamo l'annotazione in diverse banche dati derivate (Prosite, pfam etc.) per il prodotto genico.

Adesso risaliamo lungo la pagina e facciamo attenzione alle opzioni nel riquadro giallo sulla sinistra. In una box bianca al centro e' riportata la localizzazione precisa del gene sul cromosoma umano 11.

Cliccando su **Exon information** e' possibile avere informazioni piu'dettagliate sulle sequenze introniche ed esoniche. Cliccando sulle entries per ciscun esone e' possibile visualizzare quali metodi hanno permesso di identificarli come tali.

Cliccando su **Protein information** e' possibile avere la sequenza proteica e altre informazioni dettagliate sul prodotto genico (globina beta). Selezionando "Exons and SNPs" nella finestra **Show the following features** e cliccando poi su "Refresh" e' possibile visualizzare gli SNPs (Single Nucleotide Polymorphisms) noti per la stessa proteina. la sequenza proteica corrispondente al gene apparira' colorata in un codice di colori spiegato nello **schema** riportato di seguito.

Prendere nota di quale tipo di SNPs sono presenti (se sono presenti) sul gene della beta globina.

