

4. vaja: Naloge in odgovori

V04-01

Poiščite nukleotidno zaporedje za človeški jetrni protein »fatty acid binding protein 1« (FABP1). Podrobneje si oglejte tisti zapis v bazo Nucleotide, ki vsebuje celotno zaporedje mRNA/cDNA, ter identificirajte posamezne elemente zapisa (pomagajte si z <http://www.ncbi.nlm.nih.gov/Sitemap/samplerecord.html>).

Koliko nukleotidov je v kodirajoči regiji (CDS) ter koliko aminokislinskih ostankov dolgo polipeptidno verigo le-ta kodira?

Odgovor

Primer iskalnega pojma za iskanje v bazi NCBI Nucleotide (lahko uporabimo kar enostavno iskanje, saj v tem primeru dobimo že takoj relevantne rezultate): **fabp1 homo sapiens**

Eden od primernih zadetkov iskanja (ki vsebuje celotno CDS + stop kodon, kar delno razberemo že iz imena) je http://www.ncbi.nlm.nih.gov/nucore/NM_001443.2

Pozor, eden izmed zadetkov se začne s start kodonom in vsebuje zaporedje do stop kodona, brez slednjega – v tem primeru torej ne gre za (bolj ali manj) celotno zaporedje cDNA ampak samo za CDS brez STOP kodona.

Koda zapisa v bazo (zadetek s povezave zgoraj): **NM_001443** (z dodano verzijo zapisa: **NM_001443.2**)

V kodirajoči regiji (vključno s STOP kodonom) je **384 nt (=498-114; pazite, da ne odšeteje enega nukleotida preveč!!!)**, ki kodirajo **127 ak-ostankov** dolgo polipeptidno verigo ((384-3)/3).

V04-02

Poiščite še genomsko zaporedje za človeški protein FABP1. Pomagajte si s povezavo: <http://www.ncbi.nlm.nih.gov/guide/howto/obtain-genomic-sequence/>

Kako se to zaporedje razlikuje od zaporedja cDNA?

Kje v genomu (kromosom, ročica, lokus) se ta gen nahaja in koliko eksonov in intronov vsebuje?

Odgovor

Zaporedje cDNA smo že našli v prejšnjem koraku, tako da do genomskega zapisa pridemo enostavno s povezavo Gene, ki jo najdemo na strani zapisa za cDNA na desni strani (**slika 1**).

The screenshot shows the NCBI Gene database entry for FABP1 (NM_001443.2). The page is titled "Kako pridemo do zapisa za cDNA in genomskega zapisa...". The main content area displays the gene name, accession number, and a list of related sequences. A red arrow points to the "Gene" link in the "Related information" section.

| PRIMARY | REFSEQ_SPAN | PRIMARY_IDENTIFIER | PRIMARY_SPAN | COMP |
|---------|-------------|--------------------|--------------|------|
| | 1-15 | AJ45359.1 | 1-15 | |
| | 16-99 | BC02287.1 | 1-84 | |
| | 100-561 | M10617.1 | 1-442 | |
| | 562-598 | BC26155.1 | 406-442 | |

Po pregledu zapisa takoj opazimo, da je to zaporedje bistveno daljše, saj vsebuje tudi introne. Število eksonov in intronov najlažje razberemo iz grafičnega prikaza (**slika 2**)

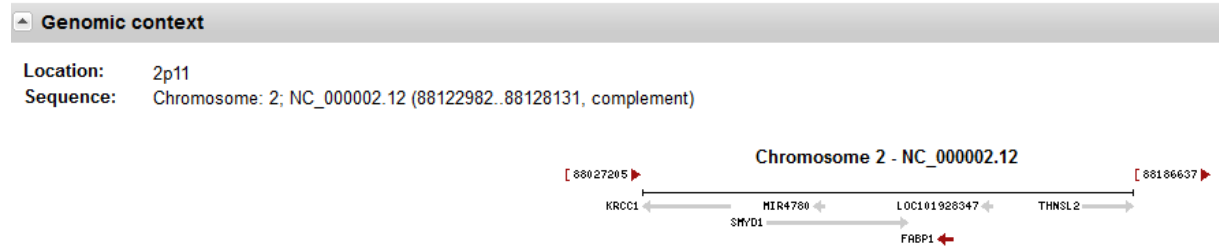
Slika 2 – prikaz intronov in eksonov



Ugotovimo, da **gen vsebuje 3 introne, skupno število eksonov je tako 4 (dva vmes in dva na koncih).**

Informacija o položaju gena na kromosomu pa je na voljo takoj v drugem razdelku (**slika 3**)

Slika 3 – lokacija na kromosomu



Razberemo, da se gen nahaja na **kromosomu 2, ročica p, lokus 11.**

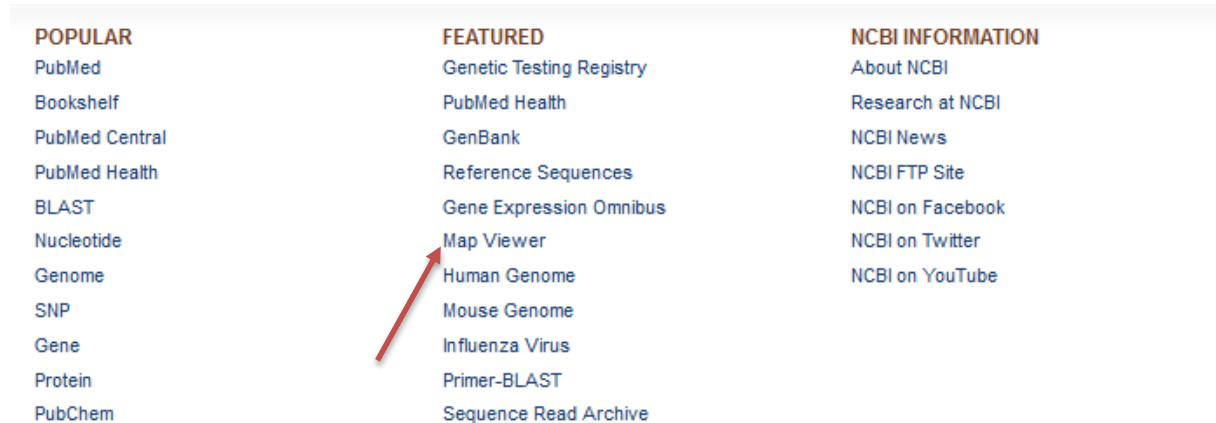
V04-03

V genomski bazi si oglejte, kateri gen se nahaja na človeškem kromosomu 7, ročica q, lokus 31.2. Kam vse vodijo povezave, nanizane pri posameznem genu?

Odgovor

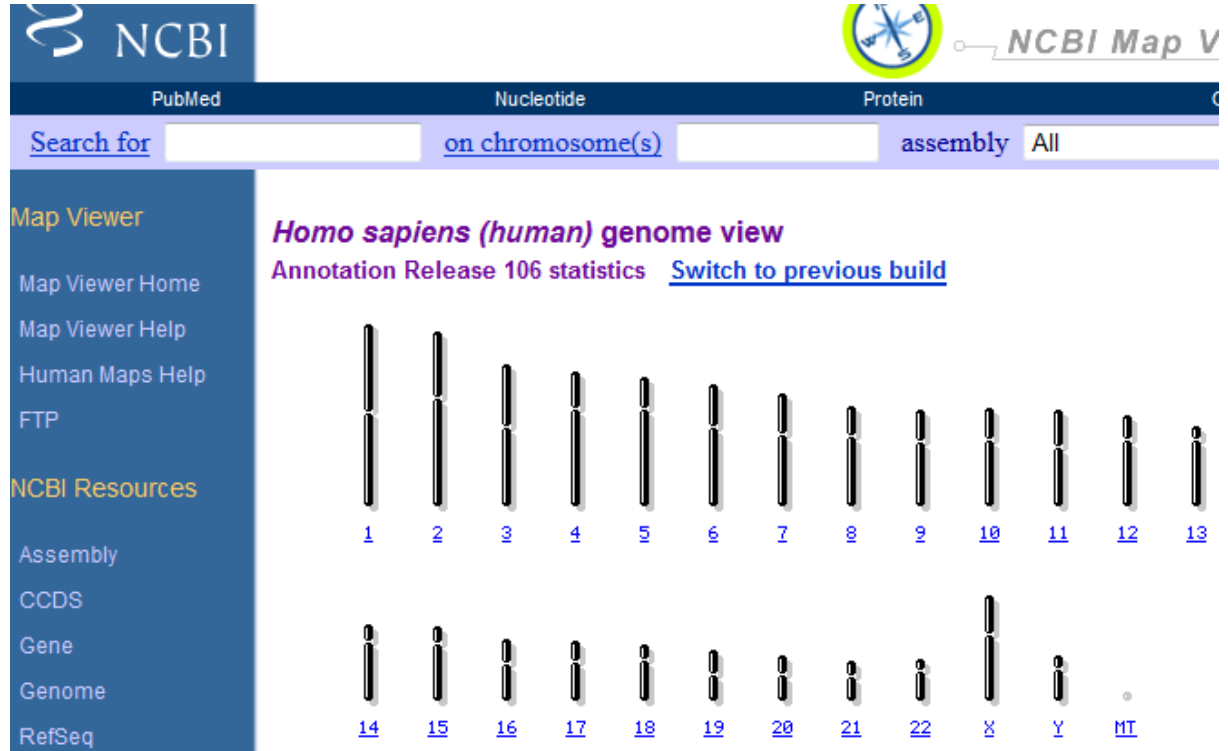
V primeru, ko poznamo lokacijo na kromosomu, ne vemo pa za kateri gen gre, je najlažji način grafično iskanje po genomskih mapah. V sklopu NCBI baze genom je temu namenjen prikaz »Map Viewer«. Pri delu z njim si lahko pomagamo z dodatnimi navodili <https://www.ncbi.nlm.nih.gov/projects/mapview/static/MapViewerHelp.html>. Do Map Viewerja lahko dostopamo že z osnovne spletne strani NCBI (<https://www.ncbi.nlm.nih.gov/>), kjer povezavo najdemo v razdelku »Featured« (slika 4)

Slika 4 – povezava do grafičnega prikaza genoma



Na naslednji strani moramo izbrati organizem »Homo Sapiens« in stisniti GO. Odpre se nam spletna stran, na kateri so narisani vsi kromosomi (slika 5)

Slika 5 - kromosomi



Zatem izberemo kromosom, po katerem želimo iskati – v našem primeru je to **kromosom 7**. Na strani, ki se nam odpre, lahko preberemo, da se na **ročici q, lokusu 31.2 nahaja gen s simbolom CFTR**. S klikom na simbol se odpre vnos za ta gen v bazi Gene. Če nas zanima kam vse vodijo dodatne povezave pri genu,

4. vaja – Baze in analiza nukleotidnih zaporedij (Miha Pavšič / marec 2014)

se lahko z miško postavimo na »Links« in se nam odpre okno, kot je to prikazano na **sliki 6**. Ne pozabite, da so lahko prikazane tudi druge povezave, kot v našem primeru - odvisno od nastavitvev.

Slika 6 – predstavitev genov na kromosomu z odprtim prikazom razlage povezav

Region Displayed: 0-159M bp

The screenshot shows a genomic browser interface with several tracks. On the left, a cytogenetic ideogram of chromosome 7 is displayed with a vertical scale from 0 to 159M bp. The main track shows gene models for various genes, including RAC1, IL6, IGFBP3, EGFR, HGF, ABCB1, PON1, CYP3A5, CYP3A4, EPO, SERPINE1, NAMPT, CAV1, MET, CFTR, LEP, BRAF, EZH2, KCNH2, and NOS3. A tooltip is open over the 'Links' column, providing a list of available database links for each gene.

| Gene Symbol | Links | Category | Description |
|-------------|----------------------------------|-------------|-------------------------------------|
| RAC1 | OMIM HGNC sv pr d | OMIM | Online Mendelian Inheritance in Man |
| IL6 | OMIM HGNC sv pr d | HGNC | HUGO Gene Nomenclature Committee |
| IGFBP3 | OMIM HGNC sv pr d | sv | Sequence Viewer |
| EGFR | OMIM HGNC sv pr d | pr | Protein sequences |
| HGF | OMIM HGNC sv pr d | dl | Sequence Download |
| ABCB1 | OMIM HGNC sv pr d | ev | Evidence Viewer |
| PON1 | OMIM HGNC sv pr d | hm | HomoloGene |
| CYP3A5 | OMIM HGNC sv pr d | sts | UniSTS |
| CYP3A4 | OMIM HGNC sv pr d | JCVI | JCVI |
| EPO | OMIM HGNC sv pr d | CCDS | Consensus CDS |
| SERPINE1 | OMIM HGNC sv pr d | SNP | SNP |
| NAMPT | OMIM HGNC sv pr dl ev hm sts SNP | best RefSeq | 7q22.3 nicotinamide pl |
| CAV1 | OMIM HGNC sv pr dl ev hm sts SNP | best RefSeq | 7q31.1 caveolin 1, cav |
| MET | OMIM HGNC sv pr dl ev hm sts SNP | best RefSeq | 7q31 OTTHUMP00 |
| CFTR | OMIM HGNC sv pr dl ev hm sts SNP | best RefSeq | 7q31.2 cystic fibrosis t |
| LEP | OMIM HGNC sv pr dl ev hm sts SNP | best RefSeq | 7q31.3 leptin |
| BRAF | OMIM HGNC sv pr dl ev hm sts SNP | best RefSeq | 7q34 v-raf murine sa |
| EZH2 | OMIM HGNC sv pr dl ev hm sts SNP | best RefSeq | 7q35-q36 enhancer of zes |
| KCNH2 | OMIM HGNC sv pr dl ev hm sts SNP | best RefSeq | 7q36.1 potassium volt |
| NOS3 | OMIM HGNC sv pr dl ev hm sts SNP | best RefSeq | 7q36 nitric oxide syn |

V04-04

Poiščite, kje (organ, tkivo, razvojna stopnja, ...) je pri človeku prisotna mRNA za nidogen. Naštejte nekaj organov/tkiv/... ter pri posameznih organih/tkivih/... dodajte kodo zapisa v bazo (»accession code«). Namig: uporabite bazo EST na NCBI. Rezultate lahko omejite z uporabo »Advanced Search«. Na kakšen način pa se danes najpogosteje določa nivo izražanja mRNA?

Odgovor

V bazi EST bomo našli več zadetkov. Vsak izmed njih ustreza mRNA, izolirani iz enega organa oz. tkiva. Točnejše informacije o organu, tkivu in razvojni stopnji najdemo znotraj vnosa v bazi. Pazimo, da gre res za nidogen 1 in da je organizem izvora človek.

Nekaj primerov zadetkov:

| Organ , tkivo | Razvojna stopnja | Accession code |
|--|-------------------------|-----------------------|
| Endotelijske celice | | DN999908.1 |
| Trebušna slinavka | | CA771392.1 |
| Trebušna slinavka | Zarodek | BU075210.1 |
| Možgani; anaplastični oligodendrogliom | | AI498817.1 |
| Trebušna slinavka | adenokarcinom | AI818389.1 |
| Jetra | | AI264298.1 |
| Maternica | Tumor | AA300269.1 |

Nivo izražanja se danes najpogosteje določa z uporabo mikromrež.

V04-05

Kakšna je razlika med EST in GSS? Poiščite GSS zapis za človeški FABP1 in si ga oglejte ter primerjajte z zapisom EST.

Odgovor

Osnovna razlika je v tem, da EST vsebuje zapise mRNA, GSS pa zapise DNA. Ko piščemo zapis za človeški FABP1 v bazi GSS takoj vidimo, da za razliko od večoh zapisov v bazi EST, v bazi GSS obstaja le en zapis za omenjeni protein (<https://www.ncbi.nlm.nih.gov/nucgss/39771898>). Vnos v bazi GSS vsebuje le zapis, ki kodira izbrani protein, ni pa nobenih podatkov o tkivu, razvojni stopnji ipd., saj vnos ne vsebuje podatkov o izražanju proteina. V bazi EST, kot smo ugotovili že v prejšnji nalogi so te podatki prisotni.

V04-06

V katero bolezen je vključen protein superoksid dismutaza 1? Kaj je pravzaprav vzrok za to bolezen? Pomagajte si z bazo OMIM (<http://www.ncbi.nlm.nih.gov/omim> ali <http://omim.org>) ter si oglejte, kaj vse zapis v to bazo vsebuje.

Odgovor

Podatke o boleznih, v katere so vključeni proteini, najlažje najdemo v bazi OMIM. V tem primeru do pravega vnosa pridemo z iskanim geslom v angleškem jeziku »superoxide dysmutase 1« (<http://omim.org/entry/147450?search=superoxide%20dysmutase%201&highlight=1%20dysmutase%20superoxide>). Zapis vsebuje alternativna imena proteina, simbol gena, lokacija gena na kromosomu, potek mapiranja, funkcija, zgodovina odkrivanja, tabela poznanih mutacij...

Iz zapisa lahko razberemo, da protein sodeluje pri amiotropni lateralni sklerozi (angl. *amyotrophic lateral sclerosis 1*). Vzrok za to bolezen še ni točno poznan, jasno pa je, da ima pomembno vlogo pri nastanku boleznih mutacije v zapisu za protein superoksid dismutaza 1

V04-07

V bazi Nucleotide poiščite cDNA zaporedje za človeški histon H3 iz družine 3B. Zaporedje v formatu FASTA shranite v tekstno datoteko. Iz zaporedja, ki ustreza kodirajoči regiji, odstranite številke, presledke, ... (orodje »DNA & Protein Sequence Cleaner«), in ga shranite v obliki formata FASTA v isto datoteko. Zabeležite si tudi kodo zapisa v bazo Nucleotide, da si boste lahko kasneje ogledali prevod v aminokislinsko zaporedje.

Odgovor

Zaporedje najdemo pod oznako **NM_005324.3**. Povezavo do FASTA zapisa najdemo na vrhu, tik pod naslovom. Da označimo kodirajočo regijo moramo najprej vklopiti funkcijo »Highlight Sequence Features« (**slika 7**) in po potrebi kliknemo še na CDS v razdelku »Features« (**slika 8**)

Slika 7 – vklop funkcije »Highlight Sequence Features«



Slika 8 – CDS

```

CDS
    /note="upstream in-frame stop codon"
    134..544
    /gene="H3F3B"
    /gene_synonym="H3.3B"
    /note="H3 histone, family 3A"
    /codon_start=1
    
```

Označen del (**slika 9**) skopiramo v program za odstranjevanje števil in presledkov, ki ga lahko najdete v spletni učilnici med povezavami, direktna povezava pa je http://www.cellbiol.com/scripts/cleaner/dna_protein_sequence_cleaner.php.

Slika 9 – označen CDS

```

ORIGIN
    1 gagcgcagag cggtttggtc gttcgttggg cggtgctggt tttcgtctcg tcgactgagg
    61 ctcttcctcg ggcagcggaa gcggcgcggc ggtcggagaa gtggcctaaa acttcggcgt
    121 tgggtgaaag aaatggccc gaaccaagca gactgctcgt aagtcaccog gtgggaaagc
    181 cccccgcaa cagctggcca cgaaagcgc caggaaaagc gtcacctata cggcggggt
    241 gaagaagcct catcgctaca ggcccgggac cgtggcgctt cgagagattc gtcgttatca
    301 gaagtcgacc gagctgctca tccggaagct gcccttcag aggttggtga gggagatcgc
    361 gcaggatttc aaaaccgacc tgaggtttca gagcgcagcc atcggtgccg tcgaggaggc
    421 tagcgaagcg tacctggtgg gtctgttoga agataccaac ctgtgtgcca tccacgctaa
    481 gagagtcacc atcatgccc aagacatcca gttggctcgc cggatacggg gagagagagc
    541 ttaagtgaag gcagttttta tggcgttttg tagtaaattc tgtaaaatac tttggtttaa
    601 tttgtgactt ttttgtaag aaattgttta taatatgttg catttgact taagtcattc
    661 catctttcac tcaggatgaa tgcgaaaagt gactgttcac agacctcagt gatgtgagca
    721 ctgttgctca ggagtgaaa gttgctaata tgcagaaggg atgggtgata cttcttgctt
    781
    
```

Vsebina datoteke je sedaj torej sledeča:

```

>gi|38373691|ref|NM_005324.3| Homo sapiens H3 histone, family 3B (H3.3B) (H3F3B), mRNA
GAGCGCAGAGCGGTTTGGTTCGTTTCGTTGGGCGGTGCTGGTTTTTCGCTCGTCTGACTGCGGCTCTTCCTCG
GGCAGCGGAAGCGGCGCGGGTTCGGAGAAGTGGCCTAAAACCTTCGGCGTTGGGTGAAAGAAAATGGCCC
GAACCAAGCAGACTGCTCGTAAGTCCACCGGTGGGAAAAGCCCCCGCAAACAGCTGGCCACGAAAGCCGC
CAGGAAAAGCGCTCCCTTACCGGCGGGGTGAAGAAGCCTCATCGCTACAGGCCCGGGACCGTGGCGCTT
CGAGAGATTTCGTCGTTATCAGAAGTCGACCGAGCTGCTCATCCGGAAGCTGCCCTTCCAGAGGTTGGTGA
GGGAGATCGCGCAGGATTTCAAACCGACCTGAGGTTTCAGAGCGCAGCCATCGGTGCGCTGCAGGAGGC
    
```

TAGCGAAGCGTACCTGGTGGGTCTGTTCTGAAGATACCAACCTGTGTGCCATCCACGCTAAGAGAGTCACC
ATCATGCCCAAAGACATCCAGTTGGCTCGCCGGATACGGGGAGAGAGAGCTTAAGTGAAGGCAGTTTTTA
TGGCGTTTTGTAGTAAATTCTGTAAAATACTTTGGTTTAATTTGTGACTTTTTTTGTAAGAAATTGTTTA
TAATATGTTGCATTTGTACTTAAGTCATTCCATCTTTCACTCAGGATGAATGCGAAAAGTACTGTTTAC
AGACCTCAGTGATGTGAGCACTGTTGCTCAGGAGTGACAAGTTGCTAATATGCAGAAGGGATGGGTGATA
CTTCTTGCCTTCTCATGATGCATGTTTCTGTATGTTAATGACTTGTGGGTAGCTATTAAGGTACTAGAGT
TGATAAATGTGTACAGGGTCCTTTTGCAATAAAAAGTGGTTATGACTTGATCCAAGTGTTTAACAATTGGG
GCTGTAAAGTCTGACCATACATCACTGTGATAGAATGTGGGCTTTTTCAAGGGTGAAGATACAAGTCTTA
ACCACAGTGTAACCTTACAGTTTCTTTAAAAAAGTAAACCTGGCAGCTATAGAATACACTATG
TGCATTTATAATAGCTATTTTATATATTGTAGTATCAACATTTTTAAATTAATGTTTTACATTCACAAG
TGGTGGGGAGTCTTGTCAATTAAGGTGTGTGTAATTTAGAGTCCAGTTGGTTTTCTTCTGACTGCACTTGT
TCTCATAGTAGTAAAATGCTATGCGCATTTATACCTTGATAAGTCCCTCATTCTACCACATGTTAACCT
CTAGCTGATAATGCAAACTAACTGGGGGATTTTATTTATAAGGGCTCTAGAAAAACGAGTTATTCAC
ACCAGCATCATCTTAACTAATCTGAACTAGTTAGTGCAGCTTTTCATTGTGTTGTGTGGTTGGTCTC
ATAACTAGGTTGAGTTTTTCTCCTCTGCTGAGGAAACAGTACCGAAGTTCTTTTTCTTGTGGCATTGTGA
TTATAAAAAGTGGTGTGGGGGAGGAGCACAAAAGTCCAGCCACTGAACCTCTGCCAATTAAGATGGTG
TTGGGTTAGGTTACATCTGGTACTGTCTGGGAAAATCATTTTTATAGAGATGGCCTTCCAAGTGGTTT
TAAAATTTACTGAAGTTTTAGGTCAATTATGTATGTTGACTAAATTTACAAATAAACTTGTATCCAA
CTAAGTGTCCAAAACCTAAATTTGAATGTAATAAGTTTTCACATGTCCATTATCTAGGTCTTGTATACT
AATGTTTTGAACTTAGATCATTTTCAAGTGTGTTTGGTGGATAAAGGAACCTTTTATTTATAAAGATACT
GTAGAAAAGCATGTGAACAGCTCTCTGCTTGAATTAAGTATGCAATAAGTGTGATTTGCAAGTGTGGGCT
AAGACAAAGTATAATTAATAAGCTTTTCAGCCCCCTCCGTTCCGTTAGTGTAGAAGCCACAGGTTGT
AGAAGTCACTTAAACTTCAAGTATGAAACCTTTCTTGTGCGATGATGGCCACTAAAGCATAGTACG
TGGATGTCAGTGAGACAGCATGAGAGCCAGCAGTCAAAAGCGTCCACGTTTGAAGTTAGCAACTGCT
TAAAGTTATGCCCTATTAATAATTTGCTTTCTCAAAAGTTTGGGTTAGTTTCAAATGTGATATTTTGGAGGG
AAGGTAAAGTAGGTATCTTTCAAGTCTGTGATAATGAGCTCCTATGAAAGGATGCAATATAATGACCCGCT
TTTCTAGAAAAGTTCATAATCAGCTCTGGAACAAGCACACTTGATTCCTCACTGTGCTTCAAGAATGAGATT
AAGATCAGATGTTGGAACGTGCTATGCTGTAGCGTGTCTGGAAAACAAGTACACAAACCTGGCTACGGTG
ATGAGTTAGCTTCTGCTTACTACCTGTGACAACCCAAGTGGGTGACACTAGTGAACCTTCTCCAGTCTGC
AGGCTGGCATAGAAGGCTCTTAGATTATATTGGGCAGCTTGTCAATCTGCCGAAGCAGTACTTGCATTTT
CACACTTGGCTTGGCACTCAACCCAGAAGGCGAAGATAGCTTTTTGGTTGTAGGCGGCTTCTGTATGGG
ATATCCCTCGTAAGGGTAAAGGAGCAGAGGCAAAGGAGAAAAGCAGAAGTTGCAGCTGATGCAGGTATC
CTATGCCCTTGTGATGAGACTAAAATAAAATTTTTGAAGTTAAAAAAGGATGCAATATAATGACCCGCT
AAAAAAGGATGCAATATAATGACCCGCT
AAAAAAGGATGCAATATAATGACCCGCT

>očiščen CDS

ATGGCCGAACCAAGCAGACTGCTCGTAAGTCCACCGGTGGGAAAGCCCCCGCAAACAGCTGGCCACGAAAGCCG
CCAGGAAAAGCGCTCCCTCTACCGGCGGGGTGAAGAAGCCTCATCGCTACAGGCCGGGACCGTGGCGCTTCGAGA
GATTCGTCGTTATCAGAAGTCGACCGAGCTGCTCATCCGGAAGCTGCCCTTCCAGAGGTTGGTGAGGGAGATCGC
GCAGGATTTCAAACCGACCTGAGGTTTCAGAGCGCAGCCATCGGTGCGCTGCAGGAGGCTAGCGAAGCGTACCT
GGTGGGTCTGTTCTGAAGATACCAACCTGTGTGCCATCCACGCTAAGAGAGTCACCATCATGCCCAAAGACATCCA
GTTGGCTCGCCGGATACGGGGAGAGAGAGCTTAA

V04-08

Celotno cDNA zaporedje za človeški histon H3 iz družine 3B preoblikujte tudi v obratno komplementarno zaporedje ter ga shranite v isto datoteko v FASTA formatu.

Odgovor

Tudi orodje za pripravo obratno komplementarnih zaporedij je na voljo med povezavami v spletni učilnici, [direktna povezava](http://www.cellbiol.com/scripts/complement/dna_sequence_reverse_complement.php) pa je http://www.cellbiol.com/scripts/complement/dna_sequence_reverse_complement.php. Že prej pripravljen zapis zgolj vstavimo v program in stisnemo »Do the Job!«. Pridobimo obratno komplementarno zaporedje:

```
TTAAGCTCTCTCTCCCCGTATCCGGCGAGCCAACTGGATGTCTTTGGGCATGATGGTGA  
CTCTTAGCGTGGATG  
GCACACAGGTTGGTATCTTCGAACAGACCCACCAGGTACGCTTCGCTAGCCTCCTGCAGCGC  
ACCGATGGCTGCGC  
TCTGAAACCTCAGGTCGGTTTTGAAATCCTGCGCGATCTCCCTCACCAACCTCTGGAAGGGC  
AGCTTCCGGATGAG  
CAGCTCGGTCGACTTCTGATAACGACGAATCTCTCGAAGCGCCACGGTCCCGGGCCTGTAGCGATGAGGCTTCTTC  
ACCCCGCCGGTAGAGGGAGCGCTTTTCCTGGCGGCTTTCGTGGCCAGCTGTTTGC  
GGGGGGCTTTCCCACCGGTGG  
ACTTACGAGCAGTCTGCTTGGTTCCGGCCAT
```

V04-09

S pomočjo orodja »ORF Finder« poiščite najdaljši bralni okvir v celotnem cDNA zaporedju za človeški histon H3 iz družine 3B. Čemu ustreza, koliko nukleotidov je dolgo in koliko aminokislinskih ostankov kodira? Iskanje ponovite na obratno komplementarnem zaporedju celotnega cDNA zaporedja in primerjajte rezultat.

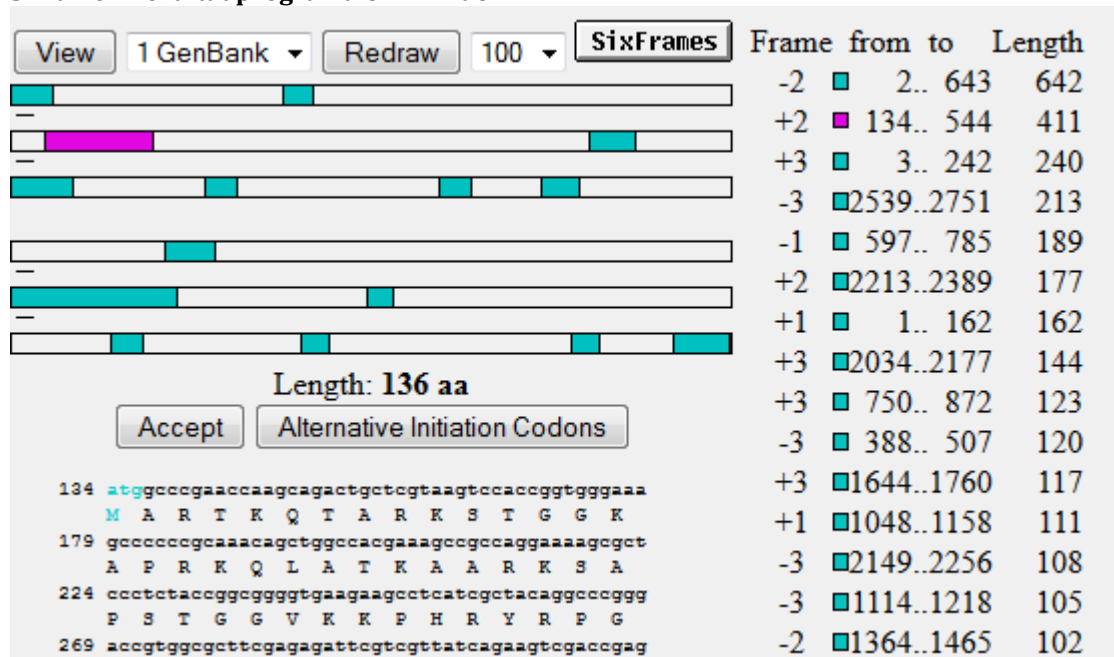
Odgovor

Zapis vstavimo v program (<https://www.ncbi.nlm.nih.gov/gorf/gorf.html>). Program sam po sebi ne naredi nič drugega kot da v vseh možnih položajih poišče zaporedja od START kodona pa do konca zapisa oz. do prvega STOP kodona.

Rezultat (slika 10) nam poda več možnih bralnih okvirjev, ki so na desni strani razvršeni po velikosti. S stiskom na obarvan kvadratik (najprej je moder, nato pa vijoličen) se nam pokaže še prevod bralnega okvirja v aminokislinsko zaporedja in dolžina le tega. Navadno je »pravilen« bralni okvir - tisti, ki se dejansko prepiše, najdaljši bralni okvir.

Vendar **POZOR**, pozorni moramo biti tudi na položaj okvirja (angl. Frame)! V našem primeru je najdaljši bralni okvir dolg 642 nukleotidov, vendar je v njegov položaj -2. Negativna številka pomeni, da gre za odprti bralni okvir v obratni smeri od tiste, ki smo jo vnesli na začetku. Kadar delamo s kodirajočimi regijami so taki rezultati nesmiselni, saj je smer prepisovanja vnaprej določena s položaji promotorjev, mest za vezavo ribosoma itd. Bralni okvir na položaju -2 med drugim niti nima nujno potrebnega STOP kodona na koncu. **Najdaljši bralni okvir, ki dejansko tudi vsebuje zapis za naš protein, je torej na položaju +2, dolg 411 nukleotidov in zapisuje za 136 aminokislinskih ostankov** (označen tudi na sliki 10).

Slika 10 – rezultat programa ORF Finder



Če uporabimo obratno komplementarno zaporedje bomo našli popolnoma enako dolge bralne okvirje (ker bodo isti), le da bo njihova lokacija na zapisu drugačna (ker je drugače številčeno) okvir pa bo imel obratno vrednost (+2 namesto -2, -1 namesto +1 itd.).

V04-10

Uporabite orodje Translate na strežniku Expasy in prevedite kodirajočo regijo cDNA zaporedja za človeški histon H3 iz družine 3B. Oglejte si možnosti, ki jih to orodje omogoča, in kakšni so rezultati.

Odgovor

Program se nahaja na povezavi <http://web.expasy.org/translate/>. Med možnostmi, ki jih imamo so trije različni načini prikaza rezultatov in izbira genskega koda, ki določa, kako naj se nukleotidno zaporedje prevede v aminokislinskega. Genskega koda, razen če natančno vemo, katerega potrebujemo, ne spreminjamo, saj bo za veliko večino zaporedij, s katerimi boste delali, standardni kod pravilen.

Kot rezultat dobimo prevode v različnih položajih bralnega okvirja. Odprti bralni okvirji, so označeni z rdečo. Najdaljši bralni okvir je na položaju +1 in dejansko obsega celotno zaporedje, njegov prevod pa je enak zaporedju proteina:

```
MARTKQTARKSTGGKAPRKQLATKAARKSAPSTGGVKKPHRYRPGTVALREIRRYQKSTE  
LLIRKLPFQRLVREIAQDFKTDLRFQSAATGALQEASEAYLVGLFEDTNLCAIHAKRVTI  
MPKDIQLARRIGERA-
```