Hands-On Seven

Beta Globin Gene and Beta Thalassemia

A) Given the coding strand (not the template strand) of a gene, do you know how to recognize the 5'UTR, 3'UTR, coding sequence (CDS), exons, and introns of the gene?
Consider the coding strand of the "beta_globin_sequence" (handout), where the "TSS: **t**ranscription **s**tart **s**ite" is underlined and the bases that are part of the CDS are in groups of 3.

a) Underline the 3 exons in blue and the 2 introns in red.
b) Underline the 5'UTR and the 3'UTR in black.
c) Underline the coding sequence (CDS) in blue
d) Draw a green rectangle around the start codon and a red rectangle around the stop codon.

B) We are going to use a gene predictor to see the exons and introns of the beta globin gene.
   • Copy the following sequence

```
>beta globin gene
tgtggagcca caccctaggg ttggccaatc tactcccagg agcaggagg gcaggagcca gggctgggca taaaagtcag
ggcagagcca tctattgctt ACATTTGCTT CTGACACAAC TGTGTTCACT AGCAACCTCA AACAGACACC
ATG GTG CAT CTG ACT CCT GAG GAG AAG TCT GCC GTT ACT GCC CTG TGG GGC AAG GTG AAC GTG GAT GAA GTT
GGT GGT GAG GCC CTG GGC AG GTTGGTATCA AGGTTACAAG ACAGGTTTAA GGAGACCAAT AGAAACTGGG CATGTGGAGA
CAGAGAAGAC TCTTGGGTTT CTGATAGGCA CTGACTCTCT CTGCCTATTG GTCTATTTTC CCACCCTTAG G CTG CTG GTG GTC
TAC CCT TGG ACC CAG AGG TTC TTT GAG TCC TTT GGG GAT CTG TCC ACT CCT GAT GCT GTT ATG GGC AAC CCT
AAG GTG AAG GCT CAT GGC AAG AAA GTG CTC GGT GCC TTT AGT GAT GGC CTG GCT CAC CTG GAC AAC CTC AAG
GGC ACC TTT GCC ACA CTG AGT GAG CTG CAC TGT GAC AAG CTG CAC GTG GAT CCT GAG AAC TTC AGG
GTGAGTCTAT GGGACGCTTG ATGTTTTCTT TCCCCTTCTT TTCTATGGTT AAGTTCATGT CATAGGAAGG GGATAAGTAA
CAGGGTACAG TTTAGAATGG GAAACAGACG AATGATTGCA TCAGTGTGGA AGTCTCAGGA TCGTTTTAGT TTCTTTTATT
TGCTGTTCAT AACAATTGTT TTCTTTTGTT TAATTCTTGC
TTTCTTTTTT TTTCTTCTCC GCAATTTTTA CTATTATACT TAATGCCTTA ACATTGTGTA TAACAAAAGG AAATATCTCT
GAGATACATT AAGTAACTTA AAAAAAAACT TTACACAGTC TGCCTAGTAC ATTACTATTT GGAATATATG TGTGCTTATT
TGCATATTCA TAATCTCCCT ACTTTATTTT CTTTTATTTT TAATTGATAC ATAATCATTA TACATATTTA TGGGTTAAAG
TGTAATGTTT TAATATGTGT ACACATATTG ACCAAATACAG GGTAATTTTG CATTTGTTAT TTTAAAAAAT GCTTTCTTCT
TTTAATATAC TTTTTTGTTT ATCTTATTTC TAATACTTTC CCTAATCTCT TTCTTTCAGG GCAATAATGA TACAATGTAT
CATGCCTCTT TGCACCATTC TAAAGAATAA CAGTGATAAT TTCTGGGTTA AGGCAATAGC AATATCTCTG CATATAAATA
TTTCTGCATA TAAATTGTAA CTGATGTAAG AGGTTTCATA TTGCTAATAG CAGCTACAAT CCAGCTACCA TTCTGCTTTT
ATTTTATGGT TGGGATAAGG CTGGATTATT CTGAGTCCAA GCTAGGCCCT TTTGCTAATC ATGTTCATAC CTCTTATCTT
CCTCCCACAG
CTC CTG GGC AAC GTG CTG GTC TGT GTG CTG GCC CAT CAC TTT GGC AAA GAA TTC ACC CCA CCA GTG CAG GCT
GCC TAT CAG AAA GTG GTG GCT GGT GTG GCT AAT GCC CTG GCC CAC AAG TAT CAC TAA GCTCGCTTTC TTGCTGTCCA
ATTTCTATTA AAGGTTCCTT TGTTCCCTAA GTCCAACTAC TAAACTGGGG GATATTATGA AGGGCCTTGA GCATCTGGAT
TCTGCCTAAT AAAAAACATT TATTTTCATT GCaatgatgt atttaaatta tttctgaata ttttactaaa aagggaatgt
gggaggtcag tgcatttaaa acataaagaa atgaagagct agttcaaacc ttgggaaaat acactatatc ttaaactcca tgaaagaa
```

   • Open a web browser and go to the Genscan server at:
      http://genes.mit.edu/GENSCAN.html

- Paste the sequence into the DNA field.
- Keep the default options and press "Run Genscan".
- Read the following explanation and answer the questions related to the table that GenScan generated.

```
Gn.Ex : gene number, exon number (for reference)
Type  : Init = Initial exon (ATG to 5' splice site)
        Intr = Internal exon (3' splice site to 5' splice site)
        Term = Terminal exon (3' splice site to stop codon)
        Sngl = Single-exon gene (ATG to stop)
        Prom = Promoter (TATA box / initation site)
        PlyA = poly-A signal (consensus: AATAAA)
S     : DNA strand (+ = input strand; - = opposite strand)
Begin : beginning of exon or signal (numbered on input strand)
End   : end point of exon or signal (numbered on input strand)
Len   : length of exon or signal (bp)
Fr    : reading frame (a forward strand codon ending at x has frame x mod 3)
Ph    : net phase of exon (exon length modulo 3)
I/Ac  : initiation signal or 3' splice site score (tenth bit units)
Do/T  : 5' splice site or termination signal score (tenth bit units)
CodRg : coding region score (tenth bit units)
P     : probability of exon (sum over all parses containing exon)
Tscr  : exon score (depends on length, I/Ac, Do/T and CodRg scores)
```

1. How many exons are predicted? _____ .
2. What do you think the last row in the prediction represents? _____ .
3. Fill in the table, giving the
   a. The start and end positions of each exon and each signal.
   b. The probability of each exon and each signal.

Exon or Signal Number	Exon or Signal Start Position	Exon or Signal End Position	Probability
1			
2			
3			
4			

C) "Changes in the Epidemiology of Thalassemia in North America: A New Minority Disease" by Elliott P. Vichinsky, et al. was published in "Pediatrics" in 2005 (issue 1, pages 1098-4275). The article can be found at:

http://pediatrics.aappublications.org/cgi/content/full/peds.2005-0843v1

The upper half of Table 1 of the article gives the mutations of the β globin gene associated with the phenotypes in the β-thalassemias in North America.

TABLE 1. β and α Mutations Associated With Each Thalassemia Phenotype

Phenotype	n
Thalassemia major	275
Codon 39 (C→T) β^0/IVS I-110 (G→A) β^+	30
IVS I-110 (G→A) β^+/IVS I-110 (G→A) β^+	28
Codon 39 (C→T) β^0/codon 39 (C→T) β^0	16
IVS I-5 (G→C) β^+/IVS I-5 (G→C) β^+	14
Codon 39 (C→T) β^0/IVS I-6 (T→C) β^+	12
IVS I-6 (T→C) β^+/IVS I-110 (G→A) β^+	11
−28 (A→G) β^+/codon 41/42 (−TCTT) β^0	10
Other genotypes	154
Thalassemia intermedia	78
IVS I-6 (T→C) β^+/IVS I-6 (T→C) β^+	5
Hb A/codon 39 (C→T) β^0	4
Codon 41/42 (−TCTT) β^0/codon 41/42 (−TCTT) β^0	4
−28 (A→G) β^+/−28 (A→G) β^+	3
Codon 39 (C→T) β^0/IVS I-6 (T→C) β^+	3
IVS I-5 (G→C) β^+/IVS I-5 (G→C) β^+	3
Codon 8/9 (+G) β^0/IVS II-1 (G→A) β^0	2
Other genotypes	54

For each of the following five mutations:

   a)   locate the mutation on the sequence,
   b)   explain what consequences on the protein it might have.

Note that start codon is in position 0.

   1) codon 6 (G→T)

   2) codon 8/9 (+G)

   3) codon 41/42 (-TCTT)

   4) codon 39 (C→T)

5) IVS-I-110 (G→A)

To learn more about beta thalassemia: http://dnalc02.cshl.edu/mydnalc/?a=dnai:1030

The following diagram gives the anatomy of an intron: