UNIVERSITY OF SKÖVDE

School of Bioscience

WRITTEN EXAMINATION

Course: Bioinformatics – Concepts and Methods

Examination: Module 5

Course code: BI760A

Credits for written examination: 1.5

Date: 13 December 2024

Examination time: 2 hours

Examination responsible: Zelmina Lubovac

Teachers concerned: Björn Olsson

Aid at the exam/appendices: None

Other

Instructions		
☐	Take a new sheet of paper for each teacher.
☒	Take a new sheet of paper when starting a new question.
☒	Write only on one side of the paper.
☒	Write your name and personal ID No. on all pages you hand in.
☒	Use page numbering.
☒	Don´t use a red pen.
☒	Mark answered questions with a cross on the cover sheet.

Grade points: 0-15 = F; 16-18 = E; 19-21 = D; 22-24 = C; 25-27 = B; 28-30 = A

Examination results should be made public within 18 working days

Good luck!

Total number of pages: 6

Question 1 (6p)

The figure below shows a region overview from Ensembl. It was created by searching for the IL6 gene (interleukin 6) in the human genome.

1a) How many transcripts (splice variants) does IL6 gene have, and how many of those transcripts are protein-coding? **(2p)**

1b) Describe what we can learn about the intron-exon structure of a transcript from the visualization in the region overview. You should pick one of the protein-coding IL6 transcripts as an example to comment on in your description. Make sure to include the accession number of the chosen transcript in your answer. **(2p)**

1c) Now compare the transcript from question 1b with another IL6 transcript from the region overview. Explain how they differ from each other. For the two chosen transcripts, you should mention, for example, which one has the largest number of exons and which one will produce the longest amino acid sequence. **(2p)**

Question 2 (6p)

For each of the following claims about Ensembl, state if the claim is true or false. You do not need to give any motivations in your answer, just writing (for each claim) "true" or "false" is sufficient.

2a) You can perform queries for specific genes, variants, and phenotypes using Ensembl's BioMart tool. **(1p)**

2b) Ensembl's genome browser allows for accessing genomic data, while BioMart is used only for displaying results from queries. **(1p)**

2c) The Ensembl Variant Effect Predictor (VEP) helps in determining the potential effects of genetic variants on gene function, including for species beyond humans. **(1p)**

2d) You can find data about single-nucleotide polymorphisms (SNPs) in Ensembl. **(1p)**

2e) Ensembl's genome browser is the only way to access the Ensembl database. **(1p)**

2f) Ensembl provides access to multiple types of entries, such as gene, transcript, and protein entries. **(1p)**

Question 3 (6p)

For each of the following claims related to gene prediction, state if the claim is true or false. You do not need to give any motivations in your answer, just writing (for each claim) "true" or "false" is sufficient.

3a) The structure of tRNA is highly variable, with no strict requirements for stem or loop length. **(1p)**

3b) *Ab initio* gene prediction relies only on the genome sequence without any additional information. **(1p)**

3c) In ORF prediction by ORF Finder tool, intermediate start codons are ignored while searching for the longest open reading frames. **(1p)**

3d) ORF Finder searches for open reading frames in only the forward reading frames (+1, +2, +3). **(1p)**

3e) ORF Finder does not distinguish between actual genes and potential coding sequences, requiring further validation. **(1p)**

3f) Homology-based gene prediction involves identifying genes in a newly sequenced genome by comparing it to related genomes. **(1p)**

Question 4 (6p)

The tRNAscan algorithm predicts tRNA genes using a rule-based approach.

4a) What does the figure below illustrate? Give example of one rule that tRNAscan applies and refer to the figure when explaining it. **(3p)**

4b) Gene prediction in prokaryote genomes is generally easier than gene prediction in eukaryote genomes. Why? **(3p)**
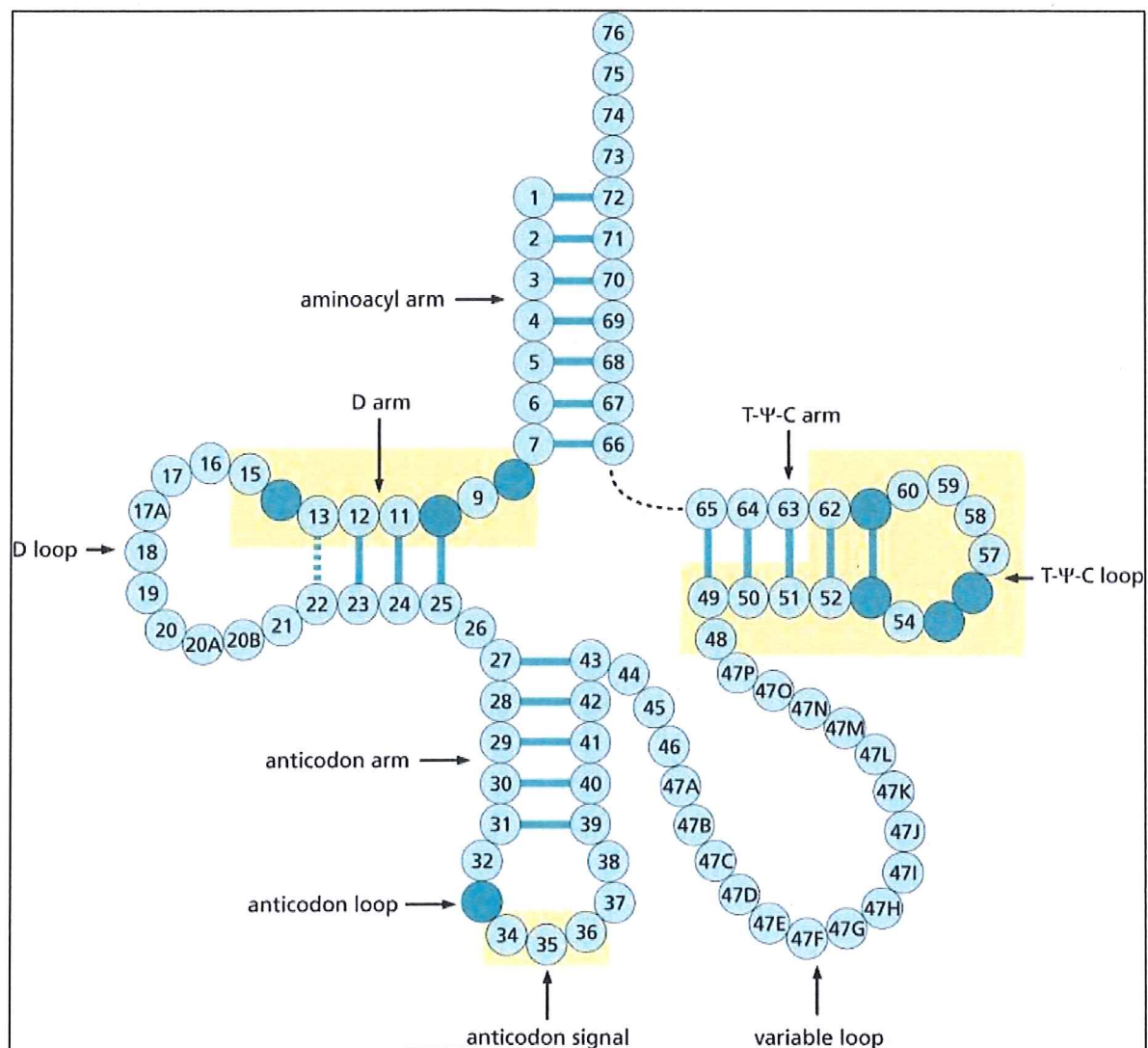
Question 5 (6p)

5a) What is a Markov model is and how can it be used to identify CpG islands in genome sequences? Support your explanation by providing an example figure. **(4p)**

5b) Provide two examples of tools that utilize Markov models (including hidden Markov models) for gene prediction. **(2p)**