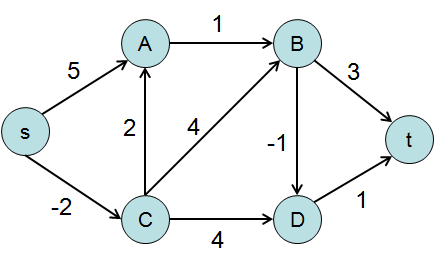
**Vipul Krishna**

**Assignment 9**

**INFO 6205 – Program Structures and Algorithms**

Question 1:

Given Graph:



1. In Bellman Ford algorithm, we can have negative edges also. First, we initialize all distances as infinity and the source as 0. We consider all edges and visit the edges (V-1) times calculating the distance from the source. Then we iterate one more time, to test the negative cycles

The step by step iteration and corresponding values of the distance are shown in the table below

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Iteration i** | **S** | **A** | **B** | **C** | **D** | **t** |
| 0 | 0 | ∞ | ∞ | ∞ | ∞ | ∞ |
| 1 | 0 | 5 | 6 | -2 | 5 | 9 |
| 2 | 0 | 0 | 2 | -2 | 2 | 3 |
| 3 | 0 | 0 | 1 | -2 | 0 | 1 |
| 4 | 0 | 0 | 1 | -2 | 0 | 1 |
| 5 | 0 | 0 | 1 | -2 | 0 | 1 |

1. Java Code Attached
2. In worst case, the number of iterations required to consider all vertices is V;

where V is the number of vertices

1. To test a negative cycle in Bellman Ford Graph, we first iterate for V-1 times and compute values of the shortest path. After iterating for one more time we get another set of solution, this shows that the graph contains a negative cycle. Thus at the end of the V iterations, if we continue finding the shortest path values, then the graph contains a negative cycle.

Question 2:

1. The paper aims to solve the shortest path problem of road maps using genetic algorithm. The solution to this problem is based on Charles Darwin theory of the natural evolution which is the key idea of Genetic Algorithms. To find the shortest path, multiple random sets are used based on the fitness values. The fitness values are calculated based on the fitness functions.

Below is the discussion which shows concrete details of genetic algorithm approach being the integral part of the presented paper.

1. Space Representation – Each town is given a unique index from 1 to N where N is the number of towns under consideration. The town index is considered as a gene of a chromosome.
2. Individual - Each town index is considered as a gene. Individuals are considered to represent a solution for the problem, i.e., individuals are paths which are representations of chromosomes. Fitness value is calculated for each individual and based on these values each individual is given a fitness rank.
3. Initial Population - The size of the population depends on the number of towns under consideration. Initially, an empty individual of size N is initialized where N is the number of towns. Individuals are generated recursively until the population contains N number of Individuals
4. Fitness function – The fitness function used for this paper is

F(x) = (1/Actual Path Length) – The # Disconnected Path

The actual path length is calculated by summing the route length of the path.

1. Rank Fitness – After the individuals run through the above function, they are assigned a rank to check their suitability and acceptance for the next cycle. This rank fitness value is updated at the end of each individual generation cycle.
2. Selection - After assigning the rank fitness, Roulette rank selection process is used where two parents. These parents are representations of individuals which are considered potential solution to the problem under discussion.
3. Cross Over - The two selected parents (individuals) undergo the cross over operation. This step is the representation of the breeding new children in the actual genetic algorithm procedure. The newly bred children would be the new set of individuals.
4. Mutation – A mutation factor of 10% is applied to the new children which are nothing but new individuals to undergo the fitness function test. The individuals are the solution to the problem which is the shortest path between two cities.

The process continues till the individuals with the best fitness rank is created. This is found out when this individual with best rank gets repeated for a longer time.

Hence, the steps adopted in this paper are essentially part of the Genetic Algorithm Concept and thus we can say that the major outline of the paper is influenced by Genetic Algorithm.

1. Java Code Attached
2. There is also a difference in the algorithm described in the paper to Genetic Algorithm. The problem in this algorithm is that according to initial selected population, the populations may at times be dominated by the individual which causes a convergence on a local maximum value. If this occurs then the algorithm would not be able to search for better solutions. Another notable difference is in the time complexity. Because of the design steps of this algorithm, the time complexity is given by the following expression

Time Complexity of Algorithm in Paper = O (V^2) for compete graphs

= O (E Log V) for normal graphs

Where V is the vertices of the graph which in this case is the towns and E is the edges.

The Time Complexity of Genetic Algorithm is given by the following expression

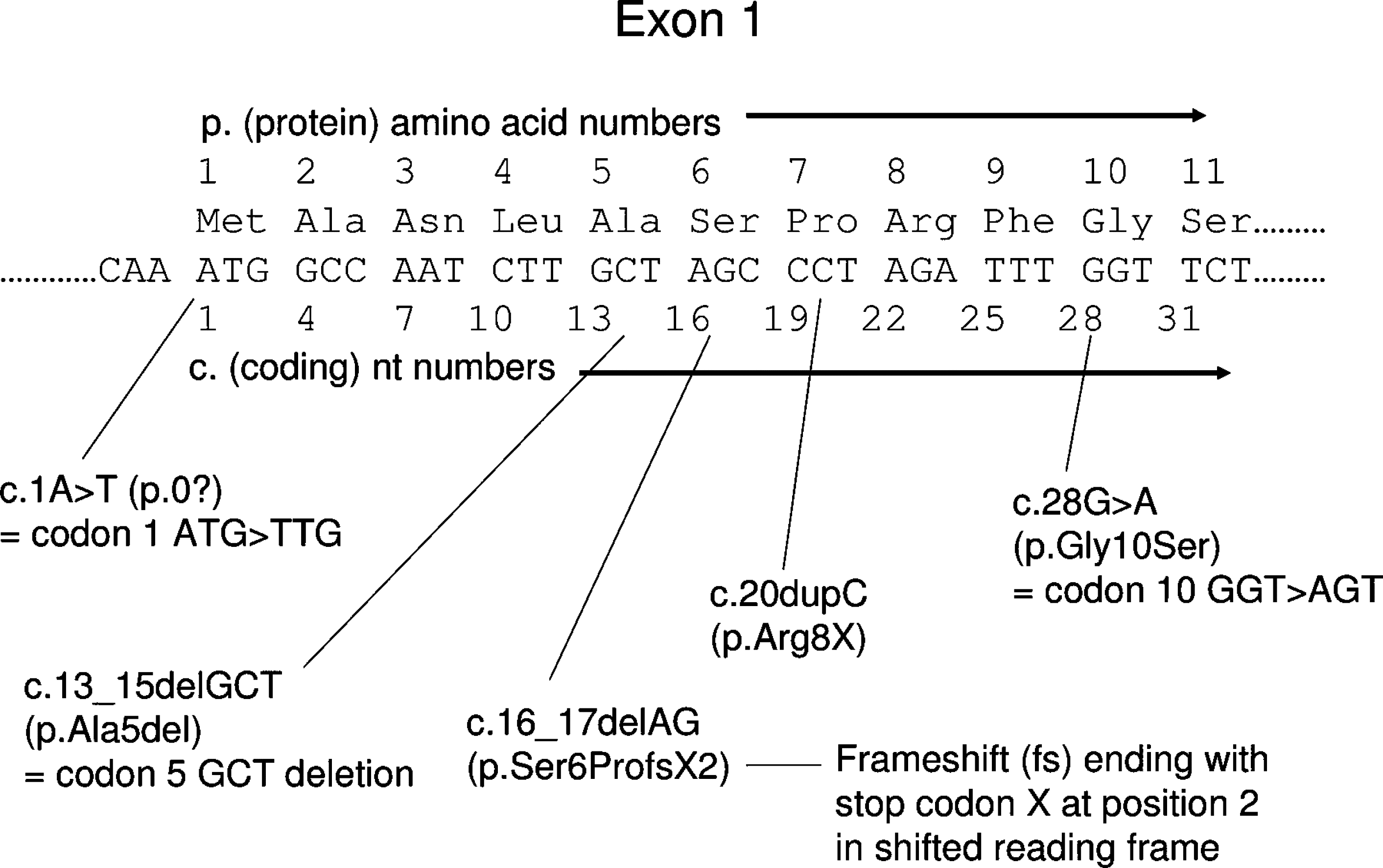
Time Complexity of Genetic Algorithm = O (g (mn))

Where m = Size of the individual and n = size of the population

Question 3

These images provide details on the standard nomenclature of Genes and their traits on Mutations.

Figure 1:



This figure is an example of a standard mutation nomenclature which is based on the coding DNA sequence.

**Terminologies Used**:

Exon – Exon is a part of the gene which encores the matured RNA information after RNA splicing.

In the figure, we see the amino acid numbered from 1 to 11. Below the numbers we have the Amino acids codes. The value of these codes is:

Met – Methionine;

Ala - Alanine

Asn - Asparagine

Leu – Leucine

Pro - Proline

Ser - Serine

Arg - Arginine

Phe - Phenylalanine

Gly – Glycine

Below the amino acids we have the nucleotide combinations which are alphabets A, C, T and G

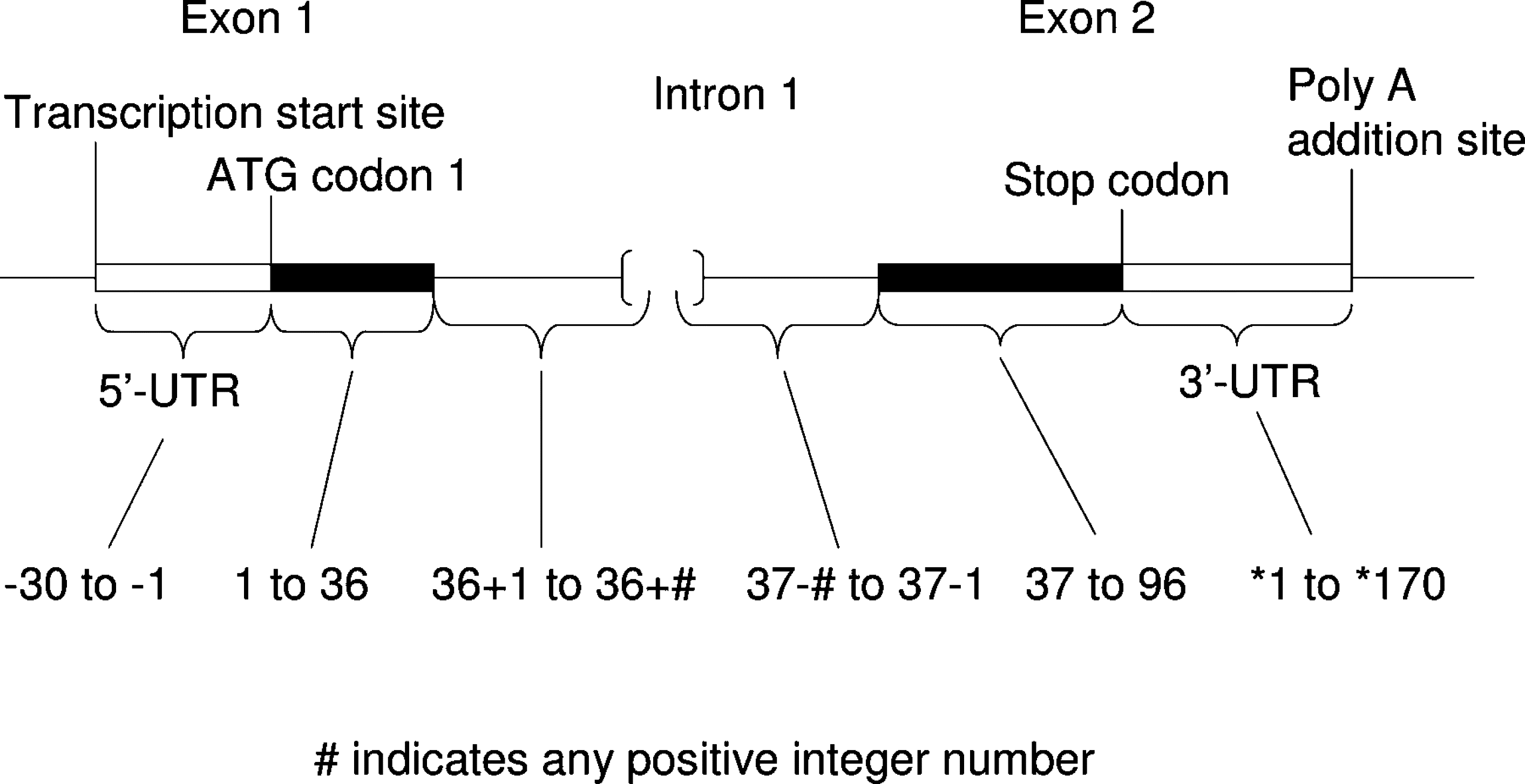
A = Adenine; C = Cytosine; T = Thymine; G = Guanine

The amino acids coding is given below that.

**Explanation**:

* c.1A>T = This denotes a change from Adenine to Thymine. Thus, the codon1 transforms from ATG to TTG. This process is described as p.0 because the changes to codon 1 are unpredictable.
* 13\_15delGCT – Codon 5 undergoes GCT deletion. Thus, the codon 5 Alanine is removed from the sequence.
* 16\_17delAG - Here, a frameshift mutation occurs and the amino acid Serine is shifted. This type of mutation shifts the way, a sequence is read.
* c.28G>A - This denotes a change of Guanine to Adenine on the 28th sequence thus making codon 10 from GGT to AGT

Figure 2:



This is an example which shows how to number nucleotides during mutation based on the standard nomenclature recommendations.

**Terminologies:**

Exon - Exon is a part of the gene which encores the matured RNA information after RNA splicing.

Intron – Introns are segments of DNA that does not participate in codes of proteins and hampers the gene formation sequence. It is a nucleotide sequence which is removed when RNA is formed.

5’-UTR – This is the 5-prime untranslated region of an mRNA upstream from the first codon.

Transcription Start Site - Transcription start site is the place where transcription begins at the end of a 5’ gene sequence.

Stop Codon - Stop codons are spots where translation into proteins terminates. This is a nucleotide triplet within the RNA. This is also known as termination codon.

3’-UTR – This is the 3-prime untranslated region that is present just after the stop codon.

Poly A Addition Site - Poly A, which signifies Polyadenylation, is the tail in a stretch of RNA sequence that consists of only adenine bases .

The last line indicates the coding sequence.

**Explanation:**

* The figure shows an example of a nucleotide numbering based on a coding DNA sequence.
* Sequences related to exons are numbered in a sequential manner from start codon to the stop codon.
* Untranslated sequences, 5’-UTR and 3’-UTR, and intronic sequences are numbered in relation to the exon sequences.