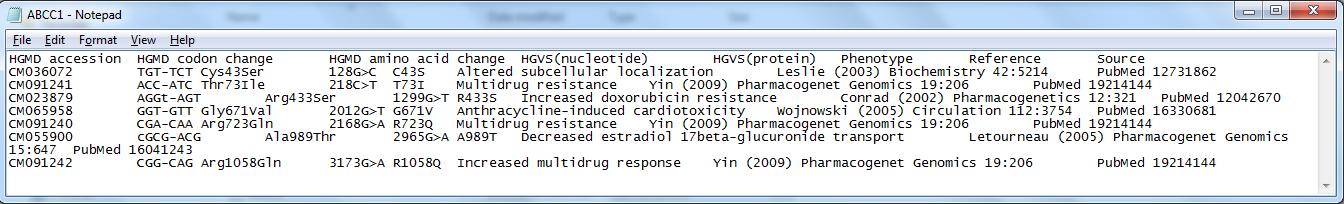
**Student Name : Point: 100**

**Problem Statement:**

Mutation analysis in terms of base mutation and amino acid mutation**.**

**Given:**

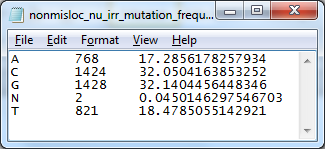
137 files containing mutation data for 137 genes. Each file contains the raw data related to mutation information for a specific gene. File name is **proteinName.txt**. Following is the snapshot of raw data.



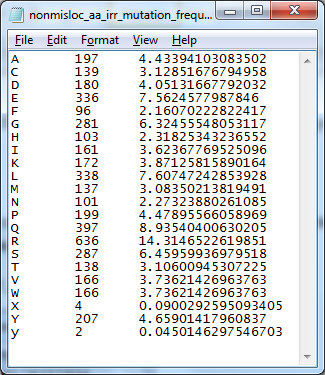
**Output:**

Two files of output. One file name is **“Final\_yourFirstName\_base\_mutation.txt”** and the other’s name is **“Final\_yourFirstName\_aa\_mutation.txt”.**

Content of **“Final\_yourFirstName\_base\_mutation.txt”** is:



Content of **“Final\_yourFirstName\_aa\_mutation.txt”** is:



As you realized that 1st column of output represents the name of bases or amino acids, 2nd column is frequency and 3rd column is %frequency. Columns are separated by a TAB.

**What to turn in:**

A zip file containing

1. Main program: **Final\_yourFirstName.pl**
2. Perl module containing all the subroutine for Final Exam: **Final\_yourFirstName.pm**
3. Perl module for reading and writing if you have used it for final exam
4. Output file1: **Final\_yourFirstName\_base\_mutation.txt**
5. Output file2: **Final\_yourFirstName\_base\_mutation.txt**

**Program to Write:**

***Special Note-1:*** *Use your already developed PERL Module for reading a file and writing to a file.*

***Special Note-2:****1st**subroutine is BONUS. If you have trouble in writing the very 1st subroutine, readFileName, you can by-pass it. The list of file name is given in file* ***filename.txt****. In that case you can start with 2nd subroutine.*

***Special Note-3:*** *Copy/type the following segments while coding:*

# \*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

# Subroutine: readFileName

#\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

#Parameter:

# Name of the directory

#Return:

# Reference to an array containing file names

#\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

sub readFileName{

#\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

#Subroutine: getMutation

#\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

#Parameter:

# Reference to an array of file names

#Return:

# Reference to an array of base mutation. Array elements are like C>A, T>G, etc

# Reference to an array of amino acid mutation. Array elements are like Q15K, R155H, etc

#\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

#\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

#Subroutine: getSample

#\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

#Parameter:

# Reference to an array of base mutation or amino acid mutation found using getMutation

#Return:

# Reference to an array of mutated base or amino acid. Array elements are 1st character of

# mutation

#\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

#***Note:*** *You need to call this subroutine twice in the main program. One call for base mutation*

*# and the other for amino acid mutation.*

#\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

#\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

#Subroutine: frequencyDistribution

#\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

#Parameter:

# Reference to an array of sample data such as mutated base, output of getSample

#Return:

# Reference to a hash of frequency distribution: key = base name, value = TAB separated

# FREQUENCY and %FREQUENCY

#\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

#***Note:*** *You need to call this subroutine twice in the main program. One call for mutated base*

*# and the other for mutated amino acid.*

#\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*