Detection and analysis of the genome for Haemoglobinopathies

*

1st Kosta Nacheski
161202

***Abstract*—Haemoglobinopathies are common monogenic disorders with diverse clinical manifestations, partly attributed to the influence of modifier genes. Haemoglobinopathies are caused by mutations in the two globin gene clusters and are characterised by a reduced or absent synthesis of globin chains in the case of the thalassaemia syndromes, mainly alpha- and beta-thalassaemia, or by defects in the haemoglobin protein structure in the case of structural haemoglobin variants, such as the Hb S that causes sickle-cell disease.In this project we will focus mainly on alpha and beta thalassemia,and sickle cell anemia.Detection of the genome of individuals with haemoglobinopathies,provided by a specific database (ex. ITHANET),isolation of the gene,mutation detection.Various methods for consideration :mutation detection in the beta-globin gene(beta thalassaemia), analysis of a Long Non-Coding RNA (lncRNA) associated with beta thalasseimia, analyzing effects of globin gene masking on gene expression analysis by RNA sequencing, determinating most common changes in the gene for synthesis on the betaglobin chains of patients with thalassaemias and abnormal hemoglobins,differences in proteins in the gene expression of haemoglobinopathies compared against with normal globin molecules.Sequencing of alpha and beta genes-detection of rare mutation responsible for alpha and beta thalassemia.As a result,we get the gene responsible for abnormal hemoglobin which causes the specific haemoglobinopathies.**

***Index Terms*—haemoglobinopathies,thalassaemia, α-thalassaemia, β-thalassaemia,sickle-cell anemia,gene,mutation**

I. Introduction

Haemoglobinopathies or Inherited hemoglobin (Hb) disorders are the most common monogenic diseases, posing a major public health problem worldwide. Haemoglobinopathies resulting from mutations in the α- or β-like globin gene clusters are the most common inherited disorders in humans, with around 7% of the world population being carriers of a globin gene mutation. Single nucleotide substitutions can lead to amino acid replacements that cause hemolytic anemias, such as sickle cell disease, or hemoglobin that are unstable or have altered oxygen affinity. Molecular defects in either regulatory or coding regions of the human α-, β- globin genes can minimally or drastically reduce their expression leading to α-,β- thalassemia, respectively. Hb is responsible for binding and transport of oxygen and carbon dioxide by red blood cells and is critical for their shape, integrity and half-life. The Hb protein complex consists of two α–like chains, and two β-like chains. HBB gene contains three coding exons separated by two introns with the size around 1600bp which has been conserved throughout the evolution process.

α-Thalassemia, one of the most common genetic diseases, is caused by deletions or point mutations affecting one to four α-globin genes. The diagnosis of α-thalassemia is complex due to a high variability of the genetic defects involved, with over 250 described mutations. With sequencing of the α- globin gene, we can use it to identify and analyze most common mutations. It has been stated that most commonly found mutations were deletions (with around 75%) and point mutations(substitution ,with 25%).

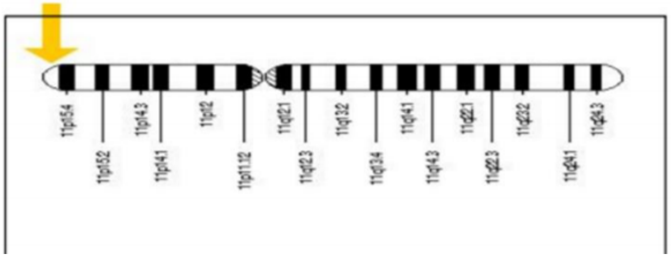
Beta thalassemia is a blood disorder that reduces the production of hemoglobin. β-Thalassemia is one of the most prevalent forms of congenital blood disorders characterized by reduced hemoglobin levels with severe complications, affecting all dimensions of life. β-thalassemia is caused by 21 important mutations out of 200 point mutations published worldwide. The type of Beta-Thalassemia depends on the severity of the mutation. If both parents carry this disorder it might be worse than if only one parent carries it. It has been reported and published that most common mutations of the HBB gene are:HBB:substitution G>C(G→C),HBB:substitution G>A(G→A),HBB:substitution G>C (G→C),HBB:substitution G>T(G→T),HBB:substitution A>C(A→C) and less but still frequent one insertion HBB : insG and also some deleterious mutations.

Sickle cell disease is caused by one particular mutation on the HBB gene, producing an abnormal version of β-globin known as haemoglobin S (HbS) which can distort red blood cells into a sickle shape. The sickle-shaped red blood cells die prematurely, which can lead to anemia. It is caused by a single point mutation in codon six of the β-globin gene from glutamic acid to valine.

II. MATERIALS AND METHODS

Haemoglobinopathies are a set of hereditary diseases caused by the abnormal structure or insufficient production of hemoglobin. Hemoglobinopathies resulting from mutations in the α- or β-like globin gene clusters are the most common inherited disorders in humans. These disorders include mutations in globin coding sequences that lead to structural changes in encoded proteins, such as the sickle hemoglobin, and mutations that alter the expression of the α- and β-globin genes such as the thalassemias.

HBB gene, which is located on chromosome 11 p15.5 , specifically located from base pair 5,225,464 to base pair 5,229,395 on chromosome 11.HBB gene contains 3 exons separated by two introns , and has a length of 1600bp. HBB protein is produced by the gene HBB which is located in the multigene locus of β-globin locus on chromosome 11,on the locus HBB specific location is located starting from the base pair 70,545,ending with 72,152. Mutations in the HBB gene are responsible for several serious haemoglobinopathies, such as sickle cell anemia and β-thalassemia.[1]



Location of the HBB gene in chromosome 11 (According to Anar Auda Ablahad - “New Approach for Analysis and Prediction of Genetic Beta-Thalassemia Mutations Based On Bioinformatics Bioedit Tools.”)

β-thalassemia is usually caused by mutations involving one or a few nucleotides in β-globin gene (HBB) or its immediate flanking regions. Mutations that completely abolish expression of the HBB are designated $\beta 0$ alleles, while other mutations in HBB cause varying degrees of quantitative reduction in β- globin expression and are classified as $\beta+$ or $\beta++$ [2] .These defects account for the vast majority of the βthalassemia alleles. They include single base substitutions, small insertions, or deletions within the gene or its immediate flanking sequences and affect almost every known stage of gene expression. They can be classified according to the mechanism by which they affect gene function: transcription, RNA processing or translation of β-globin mRNA.[3][4]

β-thalassemia is highly heterogeneous ,with around databanks of 200 mutations published worldwide.The differences in mutations differ in variation according to which population is being diagnosed. According to analyzes in the eastern regions,20 mutations that cause β - thalassemia were found to be significant,and were classified as most important.[5]

Mutation	
HBB.C.47 G>A	HBB gene,C-coding DNA ,47 location in HBB sequence >- substitution (G → A)
HBB.C.92+5 G>C	HBB gene,C-coding DNA ,92+5 location in HBB sequence >- substitution (G → C)
HBB.C.92+1 G>T	HBB gene,C-coding DNA ,92+1 location in HBB sequence >- substitution (G → T)
HBB.C.124_127 delTTCT	HBB gene,C-coding DNA ,124_127 location in HBB sequence,del - deletion (del TTCT)
HBB.C.124_127 delTTCT (Codon 41/42 - TTCT)	HBB gene,C-coding DNA ,124_127 location in HBB sequence,del - deletion (del TTCT),on Codon 41/42
HBB.C.27_28insG (Codon 8/9 (+G))	HBB gene,C-coding DNA ,27_28 location in HBB sequence,ins - insertion (insertion of base G),on Codon 8/9
HBB.C.-50 A>C	HBB gene,C-coding DNA ,-50 location in HBB sequence >- substitution (A → C)

Table 1.

The most prevalent mutation HBB.C.47 G>A, with (44.93%) ,followed by HBB.C.20 A>T (14.49%),whereas HBB.C.-50 A>C, HBB.C.92 G>C (Codon 30) mutations were the least prevalent (1.45%). Another conducted study ,grouped mutations found in β- thalassemia as transcriptional mutations,RNA processing ,RNA translation. We will list them in the tables below.

1.Transcriptional Mutations

Mutation	Type	Distribution
1)−101 (C→ T)	β^{++} (silent)	Mediterranean
2)−92 (C → T)	β^{++} (silent)	Mediterranean
3)−88 (C → T)	β^{++}	U.S. Blacks, Asian Indians
4)−87 (C→G)	β^{++}	Mediterranean
5)−87 (C→ A)	β^{++}	U.S. Blacks
6)−30 (T → A)	β^{+}	Mediterranean
7)−29 (A → G)	β^{+}	U.S. Blacks, Chinese
8)−29 (G → A)	β^{+}	Turkish
9)−28 (A → G)	β^{+}	Blacks, SE Asians
10)−27 to −26 (−AA)	β^{+}	African American

Table2

2.RNA processing

Mutation	Type	Distribution
1) IVS1-1(G→A)	β^0	Mediterranean
2) IVS1-1(G→T)	β^0	Asian Indian, SE Asian, Chinese
3) IVS1-2(T→C)	β^0	U.S. Blacks
4) IVS2-1(G→A)	β^0	Mediterranean, U.S. Blacks
5) IVS1-3' del 17 bp	β^0	Kuwaiti
6) IVS1-3' end del 25 bp	β^0	Asian Indian, UAE
7) IVS1-3'end del 44 bp	β^0	Mediterranean
8) IVS1-130 G →A	β^0	Egyptian
9) IVS2-849(A→G)	β^0	U.S. Blacks
10) IVS1-5(G→C)	β^0	Asian Indian, SE Asian, Melanesian
11) IVS1-5(G→T)	β^+	Mediterranean, N. European
12)CD27 (GCC→TCC)	β^+	Mediterranean
13) Term CD +90, del 13 bp	β^{++} (silent)	Persian

IVS- intervening sequence (intron)

Table 3

3.RNA translation

Mutation	Type	Distribution
1) **A**TG → **G**TG	β^0	Asian
2) **A**TG → **C**TG	β^0	N.European
3) A**T**G → A**C**G	β^0	South East European
4) A**T**G → A**G**G	β^0	Asian
5) A**T**G → A**A**G	β^0	N. European
6) CD6 **G**AG → **T**AG	β^0	S.American
7) CD15 T**G**G → T**A**G	β^0	Asian, Indian,
8) CD17 **A**AG → **T**AG	β^0	Asian
9) CD61 **A**AG → **T**AG	β^0	Black

Table 4,References to these mutations can be found in Forget (2001), Weatherall and Clegg (2001), and Thein and Wood (2009).

Different HBB gene mutations in Mediterranean and Middle East regions differ. Their detection rates according to the frequency and spectrum of HBB gene mutation in β-Thalassemia patients [6], concludes that the mutation HBB.c-123C>G or often HBB.c-87C>G at Mediterranean population was found to have a detection rate of 91%.While the mutation HBB.c118C>T and HBB cd39C>T in the Middle East was found to have a detection rate of 93%.

According to studies done ,from patients in Eastern Europe the most common mutations were provided to be IVS I-110 (G-A) ,HBB cd 39 (C-T) , IVS II-745 (C-G), and IVSI 1 (G-A) , but comparing to Mediterranean countries , β-Thalassemia in Eastern Europe is fairly uncommon.[7]

Sickle cell anemia, a common form of sickle cell disease, is caused by a particular mutation in the HBB gene. This mutation results in the production of an abnormal version of beta-globin called hemoglobin S or HbS. In this condition, hemoglobin S replaces both beta globin subunits in hemoglobin. The mutation changes a single amino acid in beta-globin. Detection of a single base pair mutation at 6th codon of β-globin gene is important for the diagnosis sickle cell anemia.This mutation is labeled as rs334 mutation is responsible for hemoglobin S, known as HbS, which causes sickle cell anemia.

Alpha thalassemia syndromes are caused by mutations on one or more of the four α-globin genes. Mutations could be either more commonly deletional or non-deletional. Alpha thalassemia typically results from deletions involving the HBA1 and HBA2 genes.

Databanks that provide most information about detected mutations are IthaGenes (also named ITHANET) and HbVar which has a data base of around 400 reported mutations for thalassemias,we would generally focus on detection beta thalassemia mutations and sickle cell mutation and provide a fair detection of alpha thalassemia deletion mutations .

III. Proposed approach

Haemoglobinopathies have been the central focus of many medical and biological research. Lately with the development and expansion of bioinformatics , haemoglobinopathies have taken a significant place for research , especially β- thalassemia ,which contains the most heterogeneous mutations, as such we will focus the most on in our research.

1.Related Work

Fettah A., et.al [8], Analyzed and tested for 106 Turkey patients gene sequences for B-globin gene mutations using DNA analysis. And classified as holding β-thalassemia major or β-thalassemia intermedia based on their age at diagnosis. The result showed various types of mutations types in gene regions and passed them to Turkey hospital for future gene tests.

Atanasovska B, et.al [9], Proposed approach applicable in a range of Mediterranean countries, they offered a combination of high accuracy and rapidity exploiting standard techniques and widely available equipment. Beta globin detection further adapted to particular populations by including/excluding assayed mutations. And facilitate future modifications by providing detailed information on assay design.

Verma IC, et.al [10], characterized the mutations in 1050 carriers of the beta-thalassemia gene and analyzed their regional distribution in India. The majority of beta-thalassemia carriers were migrants from Pakistan and their pattern of mutations differed from the rest. The paper result helped to successfully establish a program of genetic counselling and prenatal diagnosis of beta-thalassemia in order to reduce the burden of this disease in India.

P. Lahiry, et.al [11], suggested that “efforts to more completely characterize the HBB mutation distribution in high-risk areas, such as the Indian Subcontinent and the Middle East may lead to improved diagnosis with earlier and more effective intervention strategies. The concluded that beta-thalassemia is highly prevalent and is a major public health problem in the malaria endemic areas.

Abdul Hafeez Kandhro, et.al [12], proposed an approach of encoded and unlinked clinical laboratory data which were later evaluated with discrimination formulas, and computing approach for ensemble learning methods in bioinformatics, to get the to the statistically more correct formula for differentiation of beta-thalassemia mutations.

Monalisha Saikia Borah,et.al [13],presented a method for predicting hemoglobin variants with machine learning in bioinformatics.

2.BLAST proposed approach

The main task and aims of this proposed approach is to diagnose and detect mutations in the Beta-globin gene(HBB gene) sequence by comparisons between selected patient gene sequence with the normal gene. In order to check the normality of gene state.

3. Boyer Moore proposed approach

Furthermore ,we can implement the given sequence from the patient with a certain mutation that we got from the blast approach and implement it in a code, using the Boyer Moore algorithm. We can then use this sequence to compare it by characters with the healthy HBB gene sequence until we find a mismatch on the point where the mutation is present in the patient sequence.

4.Sickle cell anemia proposed approach

Sickle cell anemia as compared to thalassemia is caused by known particular mutation on the HBB gene.As such the approach would be slightly different.Using Biopython we can use the Entrez database to search for records of anemia present in homo sapiens sequences.From that we would obtain the features of each of the found anemia’s sequence.We can then use this sequences to be compared and analysed.

A. BIOINFORMATICS TOOLS AND PROCESSES

1.FASTA: FASTA: For every bioinformatics tools selected to complete tasks. FASTA format is required to be set as a standard expression of the entire file to the elected bioinformatics tool. FASTA file could also help in find a sequence of the required gene against keywords.

2. ClustalW: it is a powerful technique for checkup the sequences similarity whether the patient has malignant mutation related to abnormal disorder or no. by using it within BioEdit package. It can also provide the way to monitor the processes of transcription and translation to get protein sequences that control the function in gene. In this approach, ClustalW accept FASTA file that contains normal gene sequence (HBB gene sequence) and sequence from a patient HBB gene in order to compare both sequences.

3. BLAST: the basic local alignment search bioinformatics tools to searching for similarities between biological sequences which performs comparisons between pairs of sequences, searching for regions of local similarity to start sequence analysis.

The procedure will follow as , firstly we make FASTA file from HBB gene from a patient provided by the databases we use (ITHAGENES and HbVar).

Check Nucleotide alignment Of The Entire Sequences.

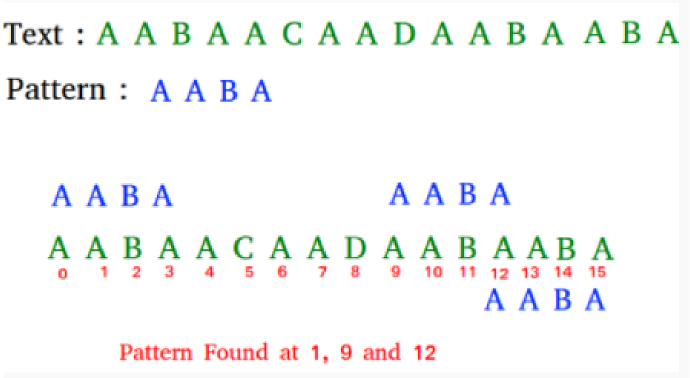
Seq1 !=Seq2?

Use Alignment Tool(Pairwise Alignment) for Similarity Check.

Differences Found? (Seq1 !=Seq2)? Then

HBB Gene is at Risk and it is Candidate for Thalassemia Disease.

4. Boyer Moore Algorithm for Pattern Searching. Boyer Moore algorithm preprocesses the pattern. Boyer Moore is a combination of following two approaches : Bad Character Heuristic and Good Suffix Heuristic. It processes the pattern and creates different arrays for both heuristics. At every step, it slides the pattern by the max of the slides suggested by the two heuristics. So it uses best of the two heuristics at every step.Unlike the previous pattern searching algorithms, Boyer Moore algorithm starts matching from the last character of the pattern.


Text : A A B A A C A A D A A B A A B A
Pattern : A A B A
A A B A
A A B A
A A B A A C A A D A A B A A B A
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
A A B A
Pattern Found at 1, 9 and 12

Example of the Booyer-Moore sequence compare)

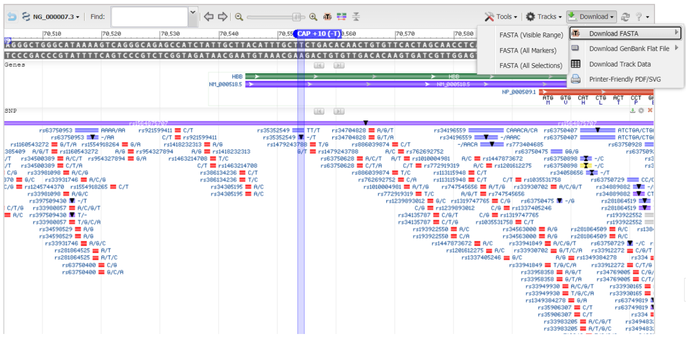
IV. Experimental Results

The Results from implementing the approach based on comparing HBB gene which is the only gene causes B-Thalassemia can be implemented in many step as:

1.Obtaining the normal HBB gene sequence from official bioinformatics databases according to the suitable environment in this system, the normal gene adopted from NCBI reference genome databases as follows:

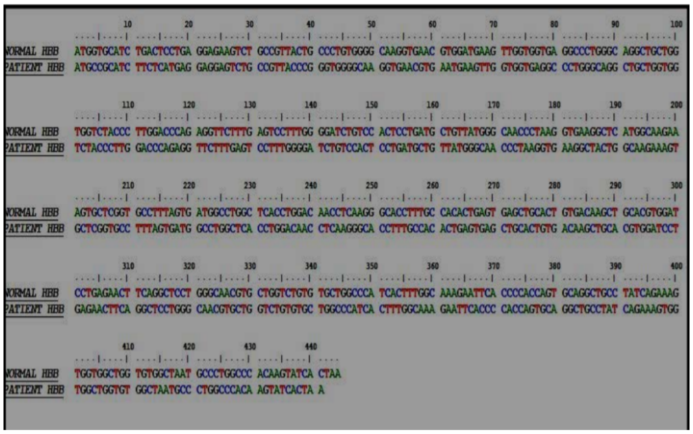
NCBI—> Gene-> Reference genome-> HBB gene sequence

2.The important stage begins in finding a provided HBB sequence with a mutation. We can use HBB gene sequence databases , IthaGenes or HbVar to find such a sequence.



(Example of Obtaining a sequence with a mutation (In this case a deletion mutation) Source:IthaGenes)

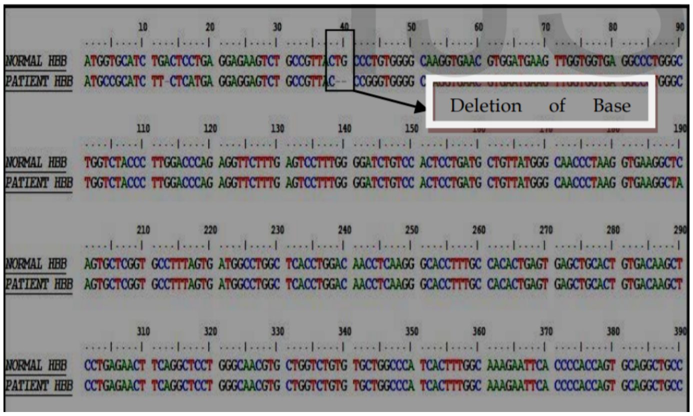
3. The important step begins in this stage after analysis and search. In this stage the normal HBB gene sequence and patient gene sequence is used as FASTA file input to the bioedit package tool to test the patient's gene if its healthy or hold somatic mutations that causes thalassemia.



```
                    10         20         30         40         50         60         70         80         90        100
NORMAL HBB   ATGGTGCATC TGACTCCTGA GGAGAAGTCT GCCGTTACTG CCCTGTGGGG CAAGGTGAAC GTGGATGAAG TTGGTGGTGA GGCCCTGGGC AGGCTGCTGG
PATIENT HBB  ATGCCGCATC TTCTCATGAG GAGGAGTCTG CCGTTACCCG GGTGGGGCAA GGTGAACGTG AATGAAGTTG GTGGTGAGGC CCTGGGCAGG CTGCTGGTGG

                   110        120        130        140        150        160        170        180        190        200
NORMAL HBB   TGGTCTACCC TTGGACCCAG AGGTTCTTTG AGTCCTTTGG GGATCTGTCC ACTCCTGATG CTGTTATGGG CAACCCTAAG GTGAAGGCTC ATGGCAAGAA
PATIENT HBB  TCTACCCTTG GACCCAGAGG TTCTTTGAGT CCTTTGGGGA TCTGTCCACT CCTGATGCTG TTATGGGCAA CCCTAAGGTG AAGGCTACTG GCAAGAAAGT

                   210        220        230        240        250        260        270        280        290        300
NORMAL HBB   AGTGCTCGGT GCCTTTAGTG ATGGCCTGGC TCACCTGGAC AACCTCAAGG GCACCTTTGC CACACTGAGT GAGCTGCACT GTGACAAGCT GCACGTGGAT
PATIENT HBB  GCTCGGTGCC TTTAGTGATG GCCTGGCTCA CCTGGACAAC CTCAAGGGCA CCTTTGCCAC ACTGAGTGAG CTGCACTGTG ACAAGCTGCA CGTGGATCCT

                   310        320        330        340        350        360        370        380        390        400
NORMAL HBB   CCTGAGAACT TCAGGCTCCT GGGCAACGTG CTGGTCTGTG TGCTGGCCCA TCACTTTGGC AAAGAATTCA CCCCACCAGT GCAGGCTGCC TATCAGAAAG
PATIENT HBB  GAGAACTTCA GGCTCCTGGG CAACGTGCTG GTCTGTGTGC TGGCCCATCA CTTTGGCAAA GAATTCACCC CACCAGTGCA GGCTGCCTAT CAGAAAGTGG

                   410        420        430        440
NORMAL HBB   TGGTGGCTGG TGTGGCTAAT GCCCTGGCCC ACAAGTATCA CTAA
PATIENT HBB  TGGCTGGTGT GGCTAATGCC CTGGCCCACA AGTATCACTA A
```

(Normal and Patients HBB Sequence File)

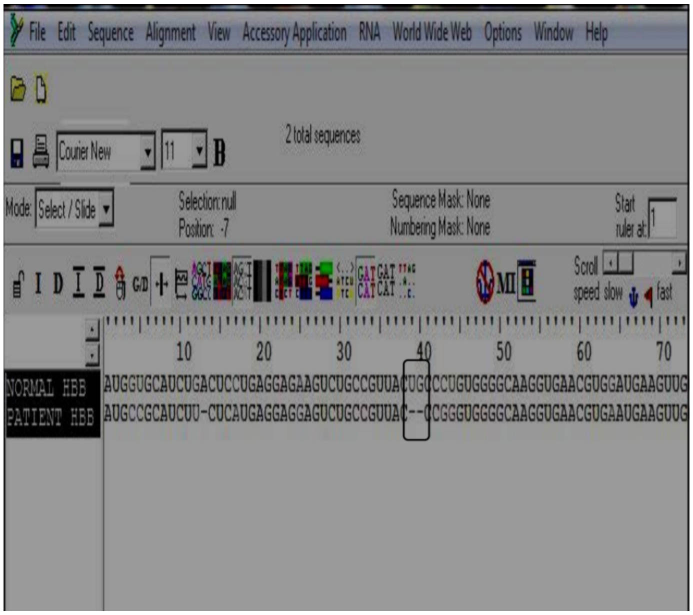
First the clustalW tool is used for alignment of the two HBB sequences at Nucleotide level.



```
                    10         20         30         40         50         60         70         80         90
NORMAL HBB   ATGGTGCATC TGACTCCTGA GGAGAAGTCT GCCGTTACTG CCCTGTGGGG CAAGGTGAAC GTGGATGAAG TTGGTGGTGA GGCCCTGGGC
PATIENT HBB  ATGCCGCATC TT-CTCATGA GGAGGAGTCT GCCGTTAC-- CCGGGTGGGG C                                    GGGC

                   110        120        130        140        150                                  190
NORMAL HBB   TGGTCTACCC TTGGACCCAG AGGTTCTTTG AGTCCTTTGG GGATCTGTCC ACTCCTGATG CTGTTATGGG CAACCCTAAG GTGAAGGCTC
PATIENT HBB  TGGTCTACCC TTGGACCCAG AGGTTCTTTG AGTCCTTTGG GGATCTGTCC ACTCCTGATG CTGTTATGGG CAACCCTAAG GTGAAGGCTA

                   210        220        230        240        250        260        270        280        290
NORMAL HBB   AGTGCTCGGT GCCTTTAGTG ATGGCCTGGC TCACCTGGAC AACCTCAAGG GCACCTTTGC CACACTGAGT GAGCTGCACT GTGACAAGCT
PATIENT HBB  AGTGCTCGGT GCCTTTAGTG ATGGCCTGGC TCACCTGGAC AACCTCAAGG GCACCTTTGC CACACTGAGT GAGCTGCACT GTGACAAGCT

                   310        320        330        340        350        360        370        380        390
NORMAL HBB   CCTGAGAACT TCAGGCTCCT GGGCAACGTG CTGGTCTGTG TGCTGGCCCA TCACTTTGGC AAAGAATTCA CCCCACCAGT GCAGGCTGCC
PATIENT HBB  CCTGAGAACT TCAGGCTCCT GGGCAACGTG CTGGTCTGTG TGCTGGCCCA TCACTTTGGC AAAGAATTCA CCCCACCAGT GCAGGCTGCC
```

(Shows the Alignment Result (Deletion Mutation))

In this stage and because the result shows mutation in some positions on patient's gene. This means that there is a chance of having Beta thalassemia. In order to make sure of this result the two sequences should convert to protein via DNA-RNA Protein sequences to find if the gene changes its function. If the protein sequence keeps the sequences similar 100% this mean that the gene is healthy, otherwise the gene would be at highly risk of Beta thalassemia disease.



Courier New 11 B

2 total sequences

Mode: Select / Slide

Selection: null
Position: -7

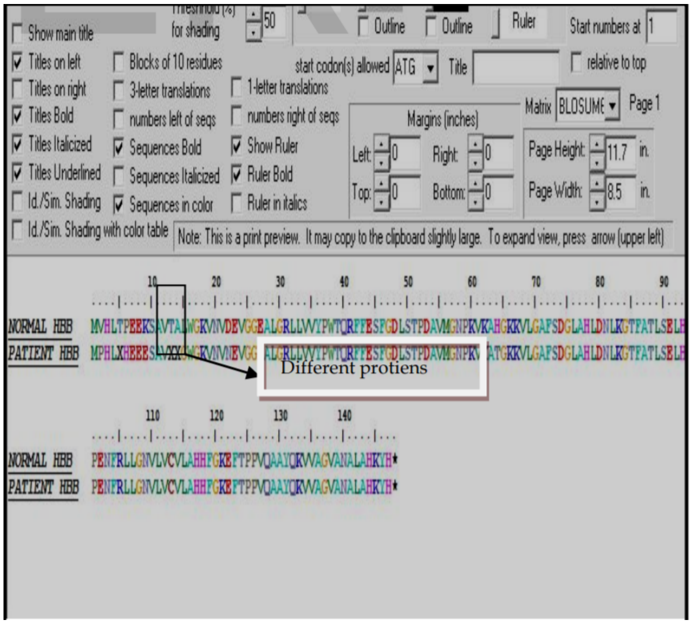
Sequence Mask: None
Numbering Mask: None

Start ruler at: 1

```
                    10        20        30        40        50        60        70
NORMAL HBB   AUGGUGCAUCUGACUCCUGAGGAGAAGUCUGCCGUUACUGCCCUGUGGGGCAAGGUGAACGUGGAUGAAGUUG
PATIENT HBB  AUGCCGCAUCUU-CUCAUGAGGAGGAGUCUGCCGUUAC--CCGGGUGGGGCAAGGUGAACGUGAAUGAAGUUG
```

(DNA → RNA (Transcription Process))

The final stage is the protein sequence convertion (translation process).If differences occure a gain then the gene is ubnormal and the disease has actually happened.



☐ Show main title
Threshold (%) for shading 50
☐ Outline ☐ Outline
Ruler
Start numbers at 1
☑ Titles on left
☐ Titles on right
☑ Titles Bold
☑ Titles Italicized
☑ Titles Underlined
☐ Id./Sim. Shading
☐ Id./Sim. Shading with color table
☐ Blocks of 10 residues
☐ 3-letter translations
☐ numbers left of seqs
☑ Sequences Bold
☐ Sequences Italicized
☑ Sequences in color
start codon(s) allowed ATG
Title
☐ relative to top
☐ 1-letter translations
☐ numbers right of seqs
☑ Show Ruler
☑ Ruler Bold
☐ Ruler in italics
Margins (inches)
Left: 0 Right: 0
Top: 0 Bottom: 0
Matrix BLOSUM6
Page 1
Page Height: 11.7 in.
Page Width: 8.5 in.

Note: This is a print preview. It may copy to the clipboard slightly large. To expand view, press arrow (upper left)

```
                     10        20        30        40        50        60        70        80        90
NORMAL HBB   MVHLTPEEKSAVTALWGKVNVDEVGGEALGRLLVVYPWTQRFFESFGDLSTPDAVMGNPKVKAHGKKVLGAFSDGLAHLDNLKGTFATLSELH
PATIENT HBB  MPHLXHEEESAVXXXWGKVNVNEVGG ALGRLLVVYPWTQRFFESFGDLSTPDAVMGNPKV ATGKKVLGAFSDGLAHLDNLKGTFATLSELH

                    110       120       130       140
NORMAL HBB   PENFRLLGNVLVCVLAHHFGKEFTPPVQAAYQKVVAGVANALAHKYH*
PATIENT HBB  PENFRLLGNVLVCVLAHHFGKEFTPPVQAAYQKVVAGVANALAHKYH*
```

(RNA → Protien Conversion (Translation Process))

V. Discussion

Our aim from the beginning was to determine an approach that could lead to detection of a mutation in a patient nucleotide sequence ,with comparison to the normal HBB gene. In the previous section we stated approaches from other researchers. The weakness of all these methods and techniques they focused in diagnostic or detecting thalassemia were based on genes mutations locally. Means that comparison of patients' gene was made to conclude specific annotations and mutations in the selected country. Without taking in consideration the abnormality of gene state whether normal or abnormal. The aims of this project we will state them as analysis, alignment and mutations check compared to a healthy gene. With two sequences necessary to fulfill the proposed test. Which compared to some other methods and techniques , Fettah A., et.al [8], tested for 106 Turkey patients gene sequences for B-globin gene mutations, Verma IC, et.al [10] characterized mutations in 1050 carriers of B-thalassemia gene and analyzed their distribution in India.In both of the mention research the approach was with used DNA(Nucleotide only).In our approach with Booyer-Moore algorithm we were able to reproduce the same thing in comparison and using the BLAST approach we've included Proteins as well , so it could be determine whether the mutation was the cause for thalassemia. Additionally since sickle cell anemia is also one of the most known haemoglobinopathies ,we've included it in our research. The results were mainly intended to be supported for Bioinformatics, although they can be useful for molecular biology of haemoglobinopathies.

VI. Conclusion

From the above information it can be well known that haemoglobinopathies are dangerous disorders which are spreading worldwide, it is not only an important public health problem but also a socio-economic problem of many countries. As thalassemia is genetically derived disorder, genetic and cellular targets are potential approaches in management of disease. With the advancement of molecular genetics technique, most of the globin chain gene variants were understood and characterized.So, it is important to take into consideration about this disorder as it may prove deadly one.

The main conclusions which obtained from implementing the proposed methods and techniques of detection and analysis are:

1. The proposed approach gives accurate results for detecting the mutations for haemoglobinopathies.

2. This new approach suggested a general prediction method based on mutational in genes that cause the disease, i.e. can implement this novel method for any disease when the mutations of its gene that caused the disease are known.

3. Offers an automatic, cost effective and friendly diagnosis system for detecting mutations for haemoglobinopathies,with previously given sequences.

Acknowledgment

The authors extend their gratitude to all of the included works for providing knowledge for the research necessary for developing this assay.

References

[1] Onda M, AkaishI J, Asaka S, Okamoto J, Miyamoto S, Mizutani K, et al. Decreased expression of haemoglobin beta (HBB) gene in anaplastic thyroid cancer and recovory of its expression inhibits cell growth. British Journal of Cancer. 2005; 92: 2216–2224. 10.1038/sj.bjc.6602634 [PMC free article] [PubMed] [CrossRef] [Google Scholar]

[2] Forough Taghavifar Mohammad Hamid Gholamreza Shariati ,"Gene expression in blood from an individual with β-thalassemia: An RNA sequence analysis"

[3] Swee Lay Thein, "The Molecular Basis of β-Thalassemia"

[4] 13. Weatherall DJ, Clegg JB (2001) Inherited haemoglobin disorders: an increasing global health problem. Bulletin of the World Health Organization 79: 704-12.

[5] Spandan Chaudhary*, Dipali Dhawan, Niraj Sojitra, Pushprajsinh Chauhan, Khyati Chandratre, Pooja S Chaudhary and Prashanth G Bagali* ,"Whole Gene Sequencing Based Screening Approach to Detect βThalassemia Mutations "

[6] Raniah S Alotibi, Eman Alharbi, Bushra Aljuhani, Bdoor Alamri, Mohieldin Elsayid, Naif M Alhawiti, Fazal Hussain, Fahad Almohareb, Cherry Colcol, Shoeb Qureshi- "The frequency and spectrum of HBB gene mutation in β-Thalassemia patients in Saudi Arabia "

[7] R. Talmaci a, J. Traeger-Synodinos b, E. Kanavakis b, D. Coriu c, D. Colita c, L. Gavrila a *-"Scanning of β-globin gene for identification of β-thalassemia mutation "

[8] Fettah A, Bayram C, Yarali N, Isik P, Kara A, Culha V, Tunc B., "Betaglobin Gene Mutations in Turkish Children with Beta-Thalassemia: Results from a Single Center Study., Mediterr J Hematol Infect Dis. 2013 Sep 2;5(1):e2013055. doi: 10.4084/MJHID.2013.055.

[9] Atanasovska B, Bozhinovski G, Plaseska-Karanfilska D, Chakalova L (2012) Efficient Detection of Mediterranean β-Thalassemia Mutations by Multiplex Single-Nucleotide Primer Extension. PLoS ONE 7(10): e48167. doi:10.1371/journal.pone.0048167

[10] Verma IC, Saxena R, Thomas E, Jain PK. "Regional distribution of beta-thalassemia mutations in India", PMID: 9225979. [PubMed - indexed for MEDLINE].

[11] P Lahiry, S A Al-Attar, R A Hegele, "Understanding BetaThalassemia with Focus on the Indian Subcontinent and the Middle East", The Open Hematology Journal, 2008, 2, 5-13.

[12] Abdul Hafeez Kandhro, MSc, Watshara Shoombuatong, PhD, Virapong Prachayasittikul, PhD, Pornlada Nuchnoi, PhD "New Bioinformatics-Based Discrimination Formulas for Differentiation of Thalassemia Traits From Iron Deficiency Anemia"

[13] Monalisha Saikia Borah, Bikram Pratim Bhuyan, Mauchumi Saikia Pathak, and P. K. Bhattacharya "Machine Learning in Predicting Hemoglobin Variants" , International Journal of Machine Learning and Computing, Vol. 8, No. 2, April 2018