IJCSE International Journal of Computer Sciences and Engineering Open Access
Research Paper Vol.-7, Issue-3, March 2019 E-ISSN: 2347-2693

Detection of the Sickle Cell Anaemia Disease by Simple and Efficient Way

Anindya Sundar De

Dept. of Information Technology, JIS College of Engineering, Kalyani, India

Corresponding Author: ani879190@gmail.com

DOI: https://doi.org/10.26438/ijcse/v7i3.6871 | Available online at: www.ijcseonline.org

Accepted: 19/Mar/2019, Published: 31/Mar/2019

Abstract— Through the proposed method it can be clearly known that a human being (specially a human baby) is suffering from sickle cell anaemia or not by a fast and efficient way. This disease is caused by mutation of the gene controlling Beta-chain of haemoglobin (Hb). It replaces Glutamic acid (GAG) present at 6th position of the Beta-chain by Valine (GTG). The mutant haemoglobin molecule undergoes polymerization under low Oxygen tension causing the change in the shape of the RBC from biconcave disc to elongated Sickle-like structure. With the help of this test, it can be known that a human is a Mutant of this disorder or not, which can save the life of new generation.

Keywords—Beta chain of Haemoglobin, Glutamic acid, Valine

I. INTRODUCTION

Bioinformatics is a branch of science which deals with biological informations and the modern computer science. It is the use of computers for the acquisition, management, and analysis of biological information. Bioinformatics is emerging and advance branch of biological science, contain Biology mathematics and Computer Science [1]. Bioinformatics, a hybrid science that links biological data with techniques for information storage, distribution, and analysis to support multiple areas of scientific research, including biomedicine. Bioinformatics is fed by high-throughput data-generating experiments, including genomic sequence determinations and measurements of gene expression patterns. Genetic information is very valuable for different disease prediction and family risk analysis. For a human DNA (Deoxyribo Nucleic Acid) there are mainly 4 nitrogen bases; they are Adenine (A), Thymine (T), Guanine (G) and Cytosine(C). In case of RNA (Ribo Nucleic Acid) there are also 4 nitrogen bases namely, Adenine (A), Uracil (U), Guanine (G) and Cytocine (C). A genetic codon has 3 nitrogen bases that is a codon is made up of 3 nitrogen bases.

DNA: Adenine(A), Thymine(T), Guanine(G), Cytosine(C)

RNA: Adenine(A), Uracil(U), Guanine(G), Cytosine(C)

Codon: ATGTACCATCGGTTAACCGGTGAGTAG

Figure 1: Codon

In the figure 1, name of the codons are ATG= "Methionine" (Start codon), TAC= "Tyrosine", CAT= "Histidine", CGG= "Arginine", TTA= "Leucine", ACC= "Threonine", GGT= "Glycine", GAG= "Glutamic acid, TAG= "Amber" (Stop codon).

A codon is a sequence of three DNA or RNA nucleotides that corresponds with a specific amino acid or stop signal during protein synthesis. DNA and RNA molecules are written in a language of four nucleotides: meanwhile, the language of proteins includes 20 amino acids. [2].

There are total 64 codons in which 61 represents amino acids and 3 codons are the stop codons (TAA, TGA, and TAG). Codons provide the key that allows these two languages to be translated into each other. Each codon corresponds to a single amino acid (or stop signal), and the full set of codons is called the genetic code [3].

The genetic code is described as degenerate, or redundant, because a single amino acid may be coded for by more than one codon. When codons are read from the nucleotide sequence, they are read in succession and do not overlap with one another. [4, 5]

Sickle cell disease is a group of disorders that affects haemoglobin, the molecule in red blood cells that delivers oxygen to cells throughout the body. People with this disorder have atypical haemoglobin molecules called haemoglobin S, which can distort red blood cells into a sickle, or crescent, shape [6]. Gene analysis is not only useful for disease prediction but also can be applied for preliminary care of the patient. [7, 8, 9]. For prediction of cardiovascular disease assay thousands of genes simultaneously using micro array is also helps a lot. [11, 12]. The sickle cell anaemia is a disease that is caused by mutation (traversion) of the gene controlling beta chain of haemoglobin. The mutated gene is called Hb(s), which causes one change in amino acid sequence of beta chain. It replaces glutamic acid (GAG) by amino acid valine (GTG). The mutant gene undergoes polymerization under low oxygen tension causing the change in the shape of the RBC from biconcave disc to elongated sickle like structure. The detection of atherosclerosis can be possible with the blood

gene prediction which has been found in recent study [13]. Clinical staging, gene expression profiling of the tumour can be to predict long-term disease recurrence and survival as well as possibly for planning treatment regimens [14, 15, 16].

The rest of the paper is organized as follows. Section II presents the proposed methodology, Section III describes the algorithmic approach i.e the step by step representation, Section IV depicts the pictorial representation i.e the flow chart. Section V gives details on data collection and results analysis. Finally Section VI contains the conclusion part.

II. PROPOSED METHODOLOGY

Through my proposed method it can be known that a human baby is suffering from Sickle cell anaemia or not in a simple, fast and efficient way. The carrier of this disease can also be traced by this method. The pictorial diagram of the method has been depicted in figure 2.

III. ALGORITHMIC APPROACH

Step 1:- Open the file in which the beta chain sequence of human haemoglobin is present, in read mode and store the data in a variable named 'hb'.

Step 2:- Convert the 'hb' string into a list

Step 3:- Print the 6th codon of haemoglobin beta chain

Step 4:- Check whether the 6th codon is 'GAG' or 'GTG' and do the following-

If hb [18] ="G" and hb [19] ="A" and hb [20] ="G"
Print "Shape of haemoglobin is not affected"
also print "The haemoglobin is Normal"

If hb [18] ="G" and hb [19] ="T" and hb [20] ="G"
Print "Shape of haemoglobin is affected"
also print "Sickle cell anaemia"

Step 5:- close the file which was opened

Step 6:- Exit

An algorithm is a finite sequence of instructions. The instructions should be precise, unambiguous and capable of being carried out in stipulated stroke finite time. The above step by step representation is the algorithmic approach of this method. This method can be represented by following the pictorial representation which describes the steps more clearly.

IV. PICTORIAL REPRESENTATION (FLOW CHART)

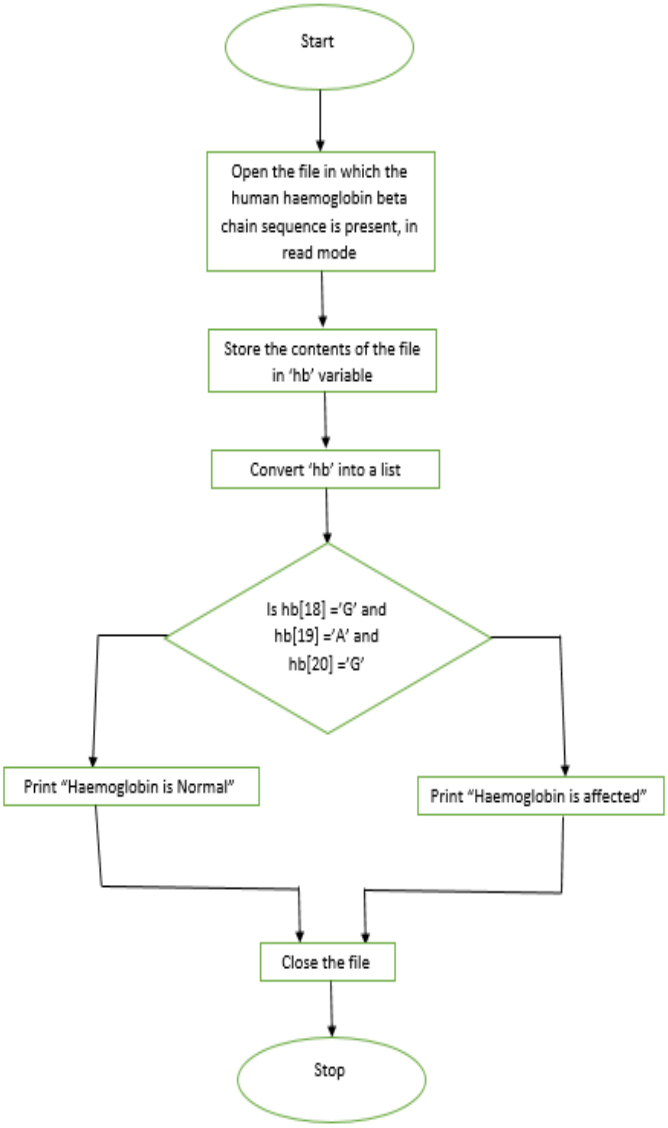

Figure 2: Pictorial Representation

V. RESULT ANALYSIS

At first the file in which the Human Hb (Haemoglobin) beta-chain sequence is present, is opened in Read mode.

ATGGTGC
ACCTGAC
TCCTGTG
GAGAAGT
CTGCCGT


```
with open("hb_n-s2.txt") as hbfile:
    hb=hbfile.read()
    hb=list(hb)

print("\nFor Normal Haemoglobin,in beta-chain\
 6th codon should be 'GAG'")
print("For Sickle shaped Haemoglobin,in beta-chain\
 6th codon should be 'GTG'")

print("Here, 6th Codon is '",hb[18],hb[19],hb[20],"'")

if(hb[18]=="G" and hb[19]=="A" and hb[20]=="G"):
        print("--> Shape of Haemoglobin is NOT affected")
        print("Result= The Haemoglobin is NORMAL")

if(hb[18]=="G" and hb[19]=="T" and hb[20]=="G"):
        print("--> Shape of Haemoglobin is AFFECTED")
        print("Result= Sickle cell anaemia")

hbfile.close()
```

Sample Results:--

1>

```
For Normal Haemoglobin,in beta-chain 6th codon should be 'GAG'
For Sickle shaped Haemoglobin,in beta-chain 6th codon should be 'GTG'
Here, 6th Codon is ' G T G '
--> Shape of Haemoglobin is AFFECTED
Result= Sickle cell anaemia
```

2>

```
For Normal Haemoglobin,in beta-chain 6th codon should be 'GAG'
For Sickle shaped Haemoglobin,in beta-chain 6th codon should be 'GTG'
Here, 6th Codon is ' G A G '
--> Shape of Haemoglobin is NOT affected
Result= The Haemoglobin is NORMAL
```

3>

```
For Normal Haemoglobin,in beta-chain 6th codon should be 'GAG'
For Sickle shaped Haemoglobin,in beta-chain 6th codon should be 'GTG'
Here, 6th Codon is ' G T G '
--> Shape of Haemoglobin is AFFECTED
Result= Sickle cell anaemia
```

4>

```
For Normal Haemoglobin,in beta-chain 6th codon should be 'GAG'
For Sickle shaped Haemoglobin,in beta-chain 6th codon should be 'GTG'
Here, 6th Codon is ' G A G '
--> Shape of Haemoglobin is NOT affected
Result= The Haemoglobin is NORMAL
```

5>

```
For Normal Haemoglobin,in beta-chain 6th codon should be 'GAG'
For Sickle shaped Haemoglobin,in beta-chain 6th codon should be 'GTG'
Here, 6th Codon is ' G A G '
--> Shape of Haemoglobin is NOT affected
Result= The Haemoglobin is NORMAL
```

6>

```
For Normal Haemoglobin,in beta-chain 6th codon should be 'GAG'
For Sickle shaped Haemoglobin,in beta-chain 6th codon should be 'GTG'
Here, 6th Codon is ' G T G '
--> Shape of Haemoglobin is AFFECTED
Result= Sickle cell anaemia
```

7>

```
For Normal Haemoglobin,in beta-chain 6th codon should be 'GAG'
For Sickle shaped Haemoglobin,in beta-chain 6th codon should be 'GTG'
Here, 6th Codon is ' G A G '
--> Shape of Haemoglobin is NOT affected
Result= The Haemoglobin is NORMAL
```

From the above sample results, we can observe that this method is able to identify that whether a human is affected from Sickle cell anaemia or not. As we can see in the sample results 1, 3 and 6, the 6^{th} codon of the human haemoglobin beta chain is 'GTG', so the haemoglobin is affected; but in case of sample result 2, 4, 5 and 7, we can see that, the 6^{th} codon of the human haemoglobin beta chain is 'GAG', so the haemoglobin is Normal.

VI. CONCLUSION

Bioinformatics is a very important field in this modern era as it is very much helpful for eradicating various types of diseases and also for finding effective drugs for human health. Sickle cell anaemia is a very crucial disease and by this method we can detect efficiently the symptoms of sickle cell anaemia in human beings.

REFERENCE

[1] Baldi P and Brunak S (1998) Bioinformatics: The Machine Learning Approach, MITPress, Cambridge, MA.
[2] Athey J, Alexaki A, Osipova E, Rostovtsev A, Santana-Quintero LV, Katneni U, Simonyan V, Kimchi-Sarfaty C (2017) A new and

updated resource for codon usage tables. BMC Bioinformatics Sept 2, 2017.
[3] Chaney JL, Clark PL (2015) Roles for Synonymous Codon Usage in Protein Biogenesis. Annu Rev Biophys, 2015 Feb 26.
[4] Hecht A, Glasgow J, Jaschke PR, Bawazer LA, Munson MS, Cochran JR, Endy D, Salit M (2017) Measurements of translation initiation from all 64 codons in E.coli : Nucleic Acids Research Apr 20, 2017.
[5] Collins FS, Green ED, Guttmacher AE, Guyer MS, U.S. National Human Genome Research Institute. A vision for the future of genomics research. Nature 2003;422:835– 47.
[6] Strouse J (2016), Sickle Cell disease. Handbook of clinical Neurology 2016; 138:311-24.
[7] 7. Roses AD. Pharmacogenetics and the practice of medicine. Nature 2000;405:857.
[8] Williams RSaG.-C PJ. The genetics of cardiovascular disease: from genotype to phenotype. Dialogues in Cardiovascular Medicine 2004; 9:3–19.
[9] Guttmacher AE, Collins FS. Genomic medicine—a primer. N Engl J Med 2002;347:1512–20.
[10] Cook SA, Rosenzweig A. DNA microarrays: implications for cardiovascular medicine. Circ Res 2002;91:559 – 64.
[11] Goldsmith ZG, Dhanasekaran N. The microrevolution: applications and impacts of microarray technology on molecular biology and medicine (review). Int J Mol Med 2004;13:483–95.
[12] Napoli C, Lerman LO, Sica V, Lerman A, Tajana G, de Nigris F. Microarray analysis: a novel research tool for cardiovascular scientists and physicians. Heart 2003;89:597– 604
[13] Patino WD, Mian OY, Kang JG, et al. Circulating transcriptome reveals markers of atherosclerosis. Proc Natl Acad SciUSA 2005;102: 3423– 8.
[14] van de Vijver MJ, He YD, van't Veer LJ, et al. A gene-expression signature as a predictor of survival in breast cancer. N Engl J Med 2002;347:1999 –2009.
[15] Bild AH, Yao G, Chang JT, et al. Oncogenic pathway signatures in human cancers as a guide to targeted therapies. Nature 2006;439: 353–7.
[16]. Berchuck A, Iversen ES, Lancaster JM, et al. Patterns of gene expression that characterize long-term survival in advanced stage serous ovarian cancers. Clin Cancer Res 2005;11:3686 –96.

