Sequence Analysis of Human HBA and HBB Thalassemia Genes


Abstract— Thalassemias are diverse group of hereditary disorders in which there is a low rate of synthesis of one or more of the globin polypeptide chains. Also, one of most common autosomal recessive disorders worldwide. Thalassemia are quantitative abnormalities of polypeptides globin chain synthesis. Normally an individual inherits two β-globin genes located one each on two chromosomes 11, and two alpha globin genes one each on two chromosome 16, from each parent i.e. normal adult hemoglobin is α2β2. The study is focused on analysis of mutated genes, their sequences and protein domains. The human alpha (HBA1) and beta (HBB) loci determine the structure of the 2 types of polypeptide chains in adult hemoglobin HbA.

Index Terms— Hereditary disorder, Autosomal, Polypeptides, Domain.

1 INTRODUCTION

Hemoglobin disorders constitute the most common lethal inherited disorders worldwide. Thalassemia is the commonest inherited hemoglobinopathy[1]. ‘Thalassaemia’ are genetic disorders and have a widespread distribution in many countries of the world. Thalassemias were among the first diseases to be characterized at the molecular level and inherited in a mendelian recessive fashion. The severe, homozygous form of the disease is called thalassaemia major, and the carrier state, in which only one defective globin gene is inherited is called the trait. Most of the thalassemias result from mutations that involve either the alpha or beta globin genes. Alpha thalassemia occurs when a gene or genes related to the alpha globin protein are missing or changed. Beta thalassemia occurs when similar gene defects affect production of the beta globin protein.

1.1 Molecular Pathogenesis

Each hemoglobin molecule contains four subunit proteins. Two of the subunit proteins are called alpha and two are called beta. Hemoglobin properly binds and releases oxygen only when two alpha subunits are connected to two beta subunits [2,3]. A pair of genes located on chromosome #16 controls the production of the alpha subunits of hemoglobin. A single gene located on chromosome #11 controls the production of the hemoglobin beta subunit. Thalassemia occurs when one or more of the genes fail to produce protein, leading to a shortage of one of the subunits.

1.2 Thalassemia Major and Minor

Individuals with thalassemia major usually present within the first two years of life with severe anemia, requiring regular red blood cell (RBC) transfusions[4,5]. Findings in untreated individuals with thalassemia major are growth retardation, pallor, jaundice, poor musculature and skeletal changes that result from expansion of the bone marrow. Carriers of beta-thalassemia are clinically asymptomatic[6]. The characteristic hematological features are microcytosis (reduced red blood cell volume), hypochromia (reduced red blood cell Hb content) and increased HbA2 level.

1.3 Beta Thalassemia
Beta-thalassemia is one of most common autosomal recessive disorders worldwide. Different molecular mechanisms like base substitutions, small deletions or insertions of one or two nucleotides in the β-globin gene are responsible for β-thalassemia [7]. Moreover, it has been found that β-thalassemia mutations are relatively population specific [8-11]. β-thalassemia is reported as the single-gene disorder.

2 MATERIAL AND METHODS

The alpha (HBA1, 141800 HBA2, 141850) and beta (HBB) loci determine the structure of the 2 types of polypeptide chains in adult hemoglobin, HbA. Mutant beta globin that sickles causes sickle cell anemia. Absence of beta chain causes beta-zero-thalassemia. Reduced amounts of detectable beta globin causes beta-plus-thalassemia.

2.1 Mutation analysis

The β-thalassemias can be caused by more than 200 different HBB gene mutations. The prevalent molecular defects are limited in each at-risk population. The most commonly used methods are reverse dot blot analysis or primer specific amplification ARMS PCR, real-time PCR or microarray technology.

2.2 Bioinformatics Approach For Sequence Analysis

Sequence analysis detects mutations in the HBB coding region and associated flanking regions. Different bioinformatics softwares were used to analyze the selected sequences.

2.3 Transcript Analysis

Transcript maps were obtained through Ensemble software. http://www.ensembl.org/Homo_sapiens/Transcript/Sequence_cDNA?db=core;g=ENSG00000206172;r=16:176680-177522;t=ENST00000320868.

3 RESULTS

The reported symbol for hemoglobin beta gene was HBB with HGNC ID: 4827. The HBB gene was located on chromosome 11. Table 1 Shows the HBB gene Symbol with other details.

Table 1

<table>
<thead>
<tr>
<th>Symbol Report: HBB</th>
</tr>
</thead>
<tbody>
<tr>
<td>APPROVED SYMBOL</td>
</tr>
<tr>
<td>APPROVED NAME</td>
</tr>
<tr>
<td>HGNC ID</td>
</tr>
<tr>
<td>PREVIOUS SYMBOLS &amp; NAMES</td>
</tr>
<tr>
<td>SYNONYMS</td>
</tr>
<tr>
<td>LOCUS TYPE</td>
</tr>
<tr>
<td>CHROMOSOMAL LOCATION</td>
</tr>
<tr>
<td>GENE FAMILY</td>
</tr>
<tr>
<td>HCOP</td>
</tr>
</tbody>
</table>

http://www.ijser.org
HBB Sequence Retrieval

**DEFINITION**  Homo sapiens hemoglobin, beta (HBB), mRNA

**ACCESSION**  NM_000518,  mRNA  626 bp

![GenBank sequence](image)

The highlighted area shows the coding region (Exon))

**HBB transcript Map analysis via Ensemble**

<table>
<thead>
<tr>
<th>Gene: HBB</th>
<th>ENSG00000244734</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Description</strong></td>
<td>hemoglobin, beta [Source: HGNC Symbol: HBB]</td>
</tr>
<tr>
<td><strong>Synonyms</strong></td>
<td>beta-globin, CD113-3, HBD</td>
</tr>
<tr>
<td><strong>Location</strong></td>
<td>Chromosome 11: 5,225,464-5,229,385 reverse strand, GRCh38:CM000673.2</td>
</tr>
<tr>
<td><strong>About this gene</strong></td>
<td>This gene has 5 transcripts (splice variants), 136 orthologues, 9 paralogues, is a member of 2 Ensembl protein families and is associated with 25 phenotypes.</td>
</tr>
<tr>
<td><strong>Transcripts</strong></td>
<td>Hide transcript table</td>
</tr>
</tbody>
</table>

**Protein domain analysis of HBB gene**
HBA Gene
The reported symbol for hemoglobin beta gene was HBA with HGNC 4823. The HBA gene was located on chromosome 16. Table 2 Shows the HBB gene Symbol with other details.

Table 2

Symbol Report: HBA1

- APPROVED SYMBOL: HBA1
- APPROVED NAME: hemoglobin, alpha 1
- HGNC ID: HGNC:4823
- PREVIOUS SYMBOLS & NAMES: -
- SYNONYMS: HBA-T3
- LOCUS TYPE: gene with protein product
- CHROMOSOMAL LOCATION: 16p13.3
- GENE FAMILY: Hemoglobin subunits
- HCOP: Orthology Predictions for HBA1
HBA Sequence Retrieval

**DEFINITION** Homo sapiens hemoglobin, alpha 1 (HBA1), mRNA.

**ACCESSION** NM_000558, mRNA 627 bp

![HBA1 mRNA sequence](image)

(The highlighted area shows the coding region (Exon))

**HBB transcript Map analysis via Ensemble**

**Gene:** HBA1 ENST00000206172

<table>
<thead>
<tr>
<th>Name</th>
<th>Transcript ID</th>
<th>bp</th>
<th>Protein</th>
<th>Biotype</th>
<th>CCDS</th>
<th>UniProt</th>
<th>RefSeq</th>
<th>Flags</th>
</tr>
</thead>
<tbody>
<tr>
<td>HBA1-001</td>
<td>ENST0000020659</td>
<td>577</td>
<td>142aa</td>
<td>Protein coding</td>
<td>CCDS103999</td>
<td>P96956</td>
<td>NM_000558</td>
<td>TSL1 GENCODE basic APPREIS P1</td>
</tr>
<tr>
<td>HBA1-006</td>
<td>ENST0000023797</td>
<td>504</td>
<td>110aa</td>
<td>Protein coding</td>
<td>-</td>
<td>G6V1N2</td>
<td>-</td>
<td>TSL2 GENCODE basic</td>
</tr>
<tr>
<td>HBA1-002</td>
<td>ENST00000247254</td>
<td>674</td>
<td>No protein</td>
<td>Retained intron</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>TSL1</td>
</tr>
<tr>
<td>HBA1-005</td>
<td>ENST00000247791</td>
<td>410</td>
<td>No protein</td>
<td>Retained intron</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>TSL2</td>
</tr>
</tbody>
</table>

**Protein domain analysis of HBB gene**

**Homo sapiens hemoglobin, alpha 1 (HBA1), mRNA**
gi|67222874|ref|NM_000558.4|

Thalassemias are a major health problem, and approximately 1 in 14 of the population are carriers for one of the subtypes. Over the past three decades, regular blood transfusions and iron chelation have dramatically improved the quality of life and transformed thalassemia from a rapidly fatal disease in early childhood to a chron-
ic disease compatible with prolonged life [12,13]. Today life expectancy varies between 25-55 years, depending on the compliance with medical treatment [54,55]. The fact that the number of patients with thalassemia decreases beyond 15 years could be explained by death mostly among children older than 15 years. This can be explained by the fact that if children are not transfused, they die before the age of 6 years and if they are transfused and non-chelated, they die before the age of 20.

5 CONCLUSION

Mutations causing the thalassemias have now been characterized and a small number of common mutations cause the bulk of disease in each particular population-base. With little more than 6 to 8 common mutations probes over 90% of thalassemic patients can now be characterized. But challenges remain in the 10% where the mutations are rare, or have not yet been determined. Newer developments in micro array technology in combination with current PCR based systems will lead to further characterisation and aid proper genetic potential identification and control of the disorder. Gene therapy is an exciting prospect. Deliveries of transgenes in stem cell based gene therapy are effective in the therapeutic management. Sequence analysis through Bioinformatics approaches shows coding regions of HBB and HBA genes. Protein sequence analysis shows various domains in each sequence. To study, sequence variations and to analyse these at molecular level can help to overcome many genetic disorders.

ACKNOWLEDGMENT

We are thankful to all colleagues for their support and fruitful suggestions.

REFERENCES

