Analysis of Human Beta Globin Gene and its Sequence Variants

Muhammad Waseem Shoaib¹, Hira Mubeen² and Shahid Raza²

1. District Head Quarter (DHQ), Hospital, Faisalabad, Pakistan
2. Lahore Garrison University, Lahore, Pakistan
*Corresponding Author’s Email: mwshoaib@hotmail.com

**ABSTRACT:** Hemoglobin is a protein commonly found in in red blood cells of all vertebrates and invertebrates. It consists of two chains, mainly alpha and beta chain. The globin genes are used as a model system to study gene expression in various eukaryotic organisms. Moreover, these genes controls the gene regulation process. Mutations in these genes can cause destruction in various regulatory pathways encoded by beta globin genes. The introns in beta globin genes mostly interrupts the sequence between codons 30-31 and between codons 104-105. The study involves a comparative genomics approach for identification and analysis of selected human beta globin gene (HBB) by using various computational tools. The nucleotide sequence of HBB gene was obtained from Genbank-NCBI and analyzed to study the transcripts within the coding region of gene by using ensemble software. These transcript maps will help to understand the regulatory features of genes, splice variants and putative indels in the reading frame. Also, the protein encoded by coding region of HBB gene was further analyzed to study various protein domains, profiles and signatures. This was done by using Interpro and CDD domain analysis software. The results are presented along with the proposed utilization of the studied gene. Comparative approaches are not only useful for evolutionary analysis but also can help to understand the function of various conserved genes.

**Keywords:** genomics, domain, transcripts, signatures

**INTRODUCTION**

Hemoglobin is the important iron containing protein in the red blood cells of all vertebrates and invertebrates. It was initially discovered in red blood cells of mammals. The hemoglobin molecule is a heterotetramer composed of two α-globin and two β-globin polypeptides in adults. It was found that each group is associated with heme group. The globin gene can be used as a model system to study gene expression in eukaryotes. Most of the studies related to thalassemia are associated with these genes. They play an important role in gene regulation and development processes. Up till now, great advances have been made for structural analysis of this gene (Borg et al., 2010; Lodish and Jacobsen, 1972).

Globin genes exists in different forms. Among all, five of them which are located on different chromosomes are well characterized. The HBA1 and HBA2 was found at
chromosomal position 16p13.3. Whereas, HBB was located at 11p15.4. All of the genes consist of at least three exons. Each exon is further separated by two introns. All introns differ in size and found closely to each other. Mostly, the genes with conserved introns can easily shuffle the exons during evolution of various proteins (Doolittle, 1972). Moreover, the introns differ in globin genes in some vertebrates which has no effect or change on vertebrate evolution (Hardison, 1998). The human β-globin gene cluster consists of five genes arranged in chromosome 11 (Lawn et al., 1978). The gene expression can be regulated particularly in a way, that alpha and beta chains can be regulated equally. Many modifications like transcriptional, posttranscriptional, and posttranslational are controlled by single β-globin gene (Smithies et al., 1978).

All genes which are involved in regulation are not active or functional every time. For activation, the transcriptional machinery will arrange on its promoter. This happens even if the gene is defective and unable to produce any protein. β-globin gene synthesis is decreased if there is any defect in β-globin gene. Studies shown expression of α and β globin genes is regulated differently in cells. A balanced production of α-globin and β-globin in erythroid cells is required for the efficient formation of hemoglobin (Weatherall and Clegg, 2008). The type of expression and time of occurrence is different in all genes encoding alpha and beta chains of globin genes. The evolutionarily pathways of globin genes have undergone multiple conversion events during mammalian evolutionary process (Song et al., 2011; Song et al., 2012). Also, the expression patterns of all globin are different for specific lineages (Hoffmann et al., 2008).

In addition, the hemoglobins produced at various stages have different affinities for oxygen and are subject to complex regulation by certain cofactors. The importance of hemoglobin in oxygen transport through blood vessels is perhaps a real support for oxidative metabolism and regulatory processes. However, the globin polypeptides can bind heme leading to successful binding of hemoglobin to oxygen by transporting it from lungs to respiring tissues.

The study involves analysis of HBB gene, its transcripts, protein domains and genetic variants found in different species. This was done by using various bioinformatics tools and software’s. The predicted domains in protein encoded by HBB gene can be further studied to analyze the 3D structure and function of protein.

**MATERIALS AND METHODS**

Several computational tools was used for identification and analysis of human HBB gene.

**Sequence Retrieval**

NCBI's is a useful data repository with tremendous sets of various genes with homologs, orthologues, protein domains and gene expression data. The nucleotide sequence of HBB gene was retrieved from NCBI-Genbank data repository and translated to get a protein sequence.
Analysis of Chromosome Map

Chromosome map was obtained for HBB gene from ensemble software. The map shows the exact chromosomal location of HBB at chromosome 11.

Transcript Identification

The transcript map of HBB gene was analyzed by using ensemble software. Transcript mapping can help to conquer genes with identified locus. The transcripts on total genomic regions are analyzed with annotated gene regions and splice variants.

Analysis of Splice Variants for Identification of Protein Domains

The transcript map for HBB gene was analyzed for identification of putative domains and protein signatures within the protein sequence. The Conserved Domain Database (CDD_NCBI) is a resource for the annotation of functional units in proteins. Study of these domain models will help to predict 3D structure and function of protein. Identification of protein domains is an important step for better understanding of protein function. The HBB proteins were analyzed by using InterPro protein domain analysis software.

RESULTS

The coding region of human HBB gene was obtained from NCBI. The complete sequence was 73308 bp in size. The coding region of gene was 1430bp in size. The highlighted area shows the coding region of HBB gene (Fig. 1).

Fig. 1: Shows the nucleotide sequence of HBB gene obtained from NCBI. The coding regions are highlighted above.
**Translated Sequence of HBB gene**

The HBB gene coding region was translated to obtain a protein sequence. The translated sequence was analyzed to check the protein encoded by particular gene. The proteins expressed by HBB gene was importantly involved in various functions with specific features and also to control different expression levels. The translated protein sequence is shown in Fig. 2.

"MVHFTAEKAAVTSKLWNSKMVNEAEAGGEALGRLLVVYPWTQRFDFSGNLSKPSAIGNPVKVAKGKVLTSGDAIKNMDNLKPAFAKLSELHCDKLHVDPENFKLGNVMVIILATHFGKEFTPEVQAWQKLVSAVAIALAHKYH"

**Fig. 2. Shows the translated protein sequence of HBB gene starting with a start codon.**

**Chromosome Map**

The chromosomal map was obtained from ensemble software. The HBB gene was located on chromosome number 11. The map is given below in Fig. 3.

**Fig. 3. Shows the chromosomal map of HBB gene located on chromosome 11.**
Analysis of human beta globin gene

Transcript Analysis

The transcripts covered by HBB gene were analyzed by using transcript analysis software. Five different types of software’s were found. Each with different size. Two of them were protein coding only. These are given in the Table 1 below.

Table 1: HBB Transcripts obtained from ensemble.

<table>
<thead>
<tr>
<th>Sr. No.</th>
<th>Name</th>
<th>Transcript</th>
<th>Protein</th>
<th>Size</th>
<th>Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>HBB-001</td>
<td>ENST00000335295</td>
<td>147aa</td>
<td>628</td>
<td>Protein Coding</td>
</tr>
<tr>
<td>2.</td>
<td>HBB-004</td>
<td>ENST00000380315</td>
<td>90aa</td>
<td>502</td>
<td>Protein Coding</td>
</tr>
<tr>
<td>3.</td>
<td>HBB-005</td>
<td>ENST00000633227</td>
<td>55aa</td>
<td>609</td>
<td>Nonsense mediated decay</td>
</tr>
<tr>
<td>4.</td>
<td>HBB-002</td>
<td>ENST00000485743</td>
<td>No Protein</td>
<td>680</td>
<td>Retained intron</td>
</tr>
<tr>
<td>5.</td>
<td>HBB-003</td>
<td>ENST00000475226</td>
<td>No Protein</td>
<td>319</td>
<td>Retained intron</td>
</tr>
</tbody>
</table>

Analysis of Splice Variants

Splice variants lying within HBB gene was analyzed. All of them was associated with different protein profiles and domains. The conserved regions was compared from different protein databases. The protein signatures and profiles were also compared which shows evolutionary relatedness. The protein profiles from three different databases namely, Pfam, PROSITE and Prints were showed maximum matches with little variation. The results are shown in Fig. 4.
Reported SNP within 100 bases of HBB gene

The genetic variations of HBB gene showed various SNP’s within the coding region of gene. All reorted SNP were having insertions/deletions. The position of contig specific to its allele was verified by contig map by using dbSNP variation viewer. The results are shown in Table 2 and Fig 5.

Table 2: SNP for rs63751076 (within 100 bases)

<table>
<thead>
<tr>
<th>Distance (Bases)</th>
<th>rs</th>
<th>Map Weight</th>
<th>Contig Accession</th>
<th>Contig Position</th>
</tr>
</thead>
<tbody>
<tr>
<td>-100</td>
<td>rs750005117</td>
<td>1</td>
<td>NT_009237.19</td>
<td>5166803</td>
</tr>
<tr>
<td>-100</td>
<td>rs755025590</td>
<td>1</td>
<td>NT_009237.19</td>
<td>5166803</td>
</tr>
<tr>
<td>-97</td>
<td>rs779006439</td>
<td>1</td>
<td>NT_009237.19</td>
<td>5166806</td>
</tr>
<tr>
<td>-96</td>
<td>rs752631303</td>
<td>1</td>
<td>NT_009237.19</td>
<td>5166807</td>
</tr>
<tr>
<td>-95</td>
<td>rs758422235</td>
<td>1</td>
<td>NT_009237.19</td>
<td>5166808</td>
</tr>
<tr>
<td>-95</td>
<td>rs768952155</td>
<td>1</td>
<td>NT_009237.19</td>
<td>5166808</td>
</tr>
<tr>
<td>-90</td>
<td>rs777373323</td>
<td>1</td>
<td>NT_009237.19</td>
<td>5166813</td>
</tr>
<tr>
<td>-89</td>
<td>rs35456885</td>
<td>1</td>
<td>NT_009237.19</td>
<td>5166814</td>
</tr>
</tbody>
</table>
Fig. 5. Shows position of rs63751076 located in intron region of HBB gene by Gene view contig annotation system.
Fig. 6. Shows the protein signatures obtained through InterPro database
Analysis of Protein Domains and Signatures

Domains are highly conserved regions of proteins. They can be used to study protein functions. Also, structure prediction can be performed by analyzing various domains. The HBB protein was analyzed for identification of conserved regions by using InterPro database. The conserved regions for globin chain of HBB protein were found with little variation at site PF00042-PS001033. The results are shown in figure 6.

DISCUSSION

Hemoglobins are commonly found in many organisms like protozoa, the cyanobacteria and the green alga (Gardner et al., 1998; Iwaasa et al., 1989; Takagi et al., 1993; Thorsteinsson et al., 1996; Kaneko et al., 1996). Similarly, family of bacterial and yeast hemoglobin shows various unique characteristics (Couture et al., 1994; Dewilde et al., 1996). The hemoglobin is a heterotetramer made up of two α-globin and two β-globin polypeptides, with an associated heme group. Both of these are encoded by the duplicated HBA1 and HBA2 genes and by the specific HBB gene. However, hemoglobins are produced only in erythroid cells, where they act as a major and essential protein. A number of different variants of hemoglobin were discovered resulting in multiple transcripts (Burmester et al., 2002; Trent and Hargrove, 2002). All behave and expression in a different fashion. Some exhibit only tissue specific expression pattern. However, the most distantly related globin found in the human genome is named as neuroglobin, encoded by NGB (Burmester et al., 2001). It was abundant in many tissues but mainly expressed in brain tissues.

Previous studies showed, the existence of five types of globin genes located on five different chromosomes: HBA1 and HBA2 at chromosomal position 16p13.3, HBB at 11p15.4, MB at 22q12.3, CYGB at 17q25.1, and NGB at 14q24.3. MB, CYGB, and NGB. These genes exists as single copy genes a compare with HBB gene which occurs in clusters with multiple genes. All of the genes consist of at least three exons separated by two introns. Presence of genes of any specific genome within its genomic region can be useful to find some flanking gene clusters for HBB gene (Flint et al., 2001; Gillemans et al., 2003). In addition, several beta globin genes were found to be linked with alpha globin genes at defined locus (Fuchs et al., 2006; Hellsten et al., 2010).

Up till now, more than 730 variants of the β-globin chain have been characterized (Weatherall and Clegg, 2001). These variants exhibits range of different types of mutations not only limited to point mutations but also includes large deletions. Such mutations can affects synthesis of HBB gene, its transcription, translation and protein structure (Patinos et al., 2004). The levels of expression of globin gene in different reactions controlled by promoters is found very low. The promoters regulates the activity of HBB gene in humans and other organisms. The activity depends upon level of regulation and modifications. A number of different mutations were detected in HBB gene. These mutations are categorized as insertions, deletions and frameshifts. A bioinformatics approach for sequence analysis can be used.
for identification of genetic variations with mutations.

CONCLUSION

The HBB gene analysis showed various functions occurring at different times during regulation and process of development. The gene transcripts showed different patterns, signatures and protein profiles. All profiles covers the maximum area of HBB gene with number of introns and splice variants. The analysis predicted several conserved regions with integrated signatures of globin chain. These regions are covered and enriched with various beta globin chains. The transcripts of HBB gene found in different organism were not all protein coding. In spite of the fact, only few of them were protein coding. Hence, analysis of HBB gene performed by these computational tools was found useful to study the gene expression and regulation process. Also, to study the globin chain at macro level for 3D structure prediction of hemoglobin beta type genes.

REFERENCES


Analysis of human beta globin gene


