

## Gene Annotation Assignment

The National Centre for Biotechnology Information (NCBI) website was used to identify the accession number of the gene, ENSG00000244734, which was found to be **NC\_000011.10**. The Open Reading Frame (ORF) Finder at the NCBI website was used to identify the ORF for this gene using the accession number. The first ORF obtained is as shown in Figure 1. This ORF is present on the minus or the antisense strand, which means that its RNA transcript is reverse complementary to the genomic sequence. The frame of this ORF is 1, which means that the first possible reading frame is transcribed and translated to give the RNA and protein products respectively.

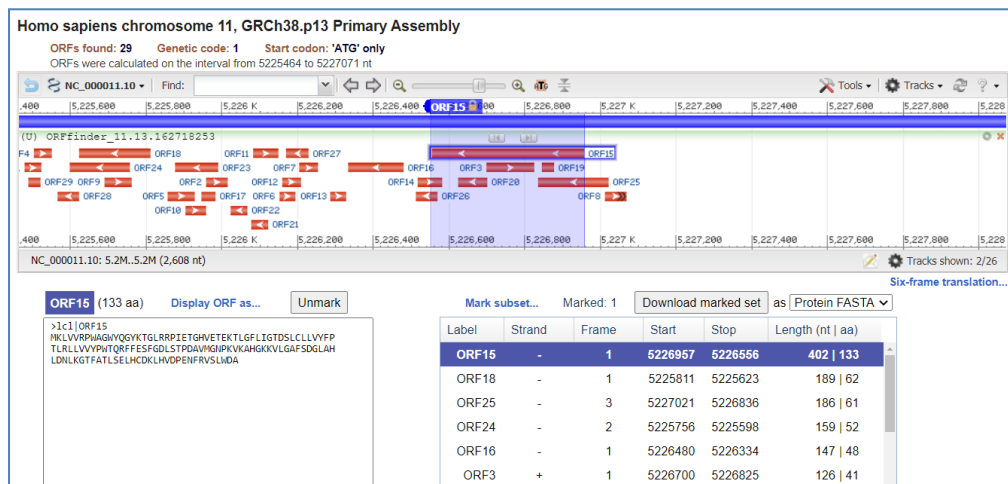


Figure 1: Identified ORFs for the unknown gene using ORF Finder at NCBI

The accession number of the unknown gene was used to perform a nucleotide alignment using the NCBI BLASTN tool to identify homologous sequences. The results of this search can be seen in Figure 2. As is evident from the figure, the query sequence shows a 100% sequence similarity with the hemoglobin subunit beta (HBB) gene in *Homo sapiens*. The maximum possible score is 2970, and the total score obtained for each of the top 3 hits is 2970. The percent identity in all 3 hits is 100% and the E value is 0, which means that the possibility of the match being a chance alignment is 0.

Job Title **NC\_000011:Homo sapiens chromosome 11, GRCh38.p13...**

RID [UWURYQK013](#) Search expires on 11-14 18:16 pm [Download All](#) ▾

Program **BLASTN** [Citation](#) ▾

Database **nt** [See details](#) ▾

Query ID [NC\\_000011.10](#)

Description **Homo sapiens chromosome 11, GRCh38.p13 Primary Assembly**

Molecule type **dna**

Query Length **1608**

Other reports [Distance tree of results](#) [MSA viewer](#) [?](#)

**Filter Results**

Organism *only top 20 will appear*  exclude

Type common name, binomial, taxid or group name

[+ Add organism](#)

Percent Identity  to  E value  to  Query Coverage  to

[Filter](#) [Reset](#)

**Descriptions** | Graphic Summary | Alignments | Taxonomy

**Sequences producing significant alignments** [Download](#) ▾ [Manage columns](#) ▾ Show **100** ▾ [?](#)

select all 100 sequences selected

	Description	Max Score	Total Score	Query Cover	E value	Per Ident	Accession
<input checked="" type="checkbox"/>	<a href="#">Homo sapiens voucher Yoruba_13_0 hemoglobin subunit beta (HBB) gene, complete cds</a>	2970	2970	100%	0.0	100.00%	<a href="#">MK476491.1</a>
<input checked="" type="checkbox"/>	<a href="#">Homo sapiens voucher Yoruba_9_0 hemoglobin subunit beta (HBB) gene, complete cds</a>	2970	2970	100%	0.0	100.00%	<a href="#">MK476483.1</a>
<input checked="" type="checkbox"/>	<a href="#">Homo sapiens voucher Mandenka_19_1 hemoglobin subunit beta (HBB) gene, complete cds</a>	2970	2970	100%	0.0	100.00%	<a href="#">MK476464.1</a>

Figure 2: Results of BLASTN search using the gene's accession number

Figure 3 shows the details of the first alignment match with the query sequence. The size of the sequence is 1608, and the identity is 100% indicating a complete match with all nucleotides. The Expect value is 0, thereby indicating that the possibility of this match being a chance alignment is 0.

**Homo sapiens voucher Yoruba\_13\_0 hemoglobin subunit beta (HBB) gene, complete cds**

Sequence ID: [MK476491.1](#) Length: 1824 Number of Matches: 1

Range 1: 88 to 1695 [GenBank](#) [Graphics](#) [Next Match](#) [Previous Match](#)

Score	Expect	Identities	Gaps	Strand
2970 bits(1608)	0.0	1608/1608(100%)	0/1608(0%)	Plus/Minus
Query 5225464	TTGCAATGAAAAATAATGTTTTTTATTAGGCAGAAATCCAGATGCTCAAGGCCCTTCATAA	5225523		
Sbjct 1695	TTGCAATGAAAAATAATGTTTTTTATTAGGCAGAAATCCAGATGCTCAAGGCCCTTCATAA	1636		
Query 5225524	TATCCCCCAGTTAGTAGTTGGACTTAGGGAACAAAGGAACCTTTAATAGAAATTGGACA	5225583		
Sbjct 1635	TATCCCCCAGTTAGTAGTTGGACTTAGGGAACAAAGGAACCTTTAATAGAAATTGGACA	1576		
Query 5225584	GCAAGAAAGCGAGCTTAGTGATACTTGTGGGCCAGGGCATTAGCCACACAGCCACCCACT	5225643		
Sbjct 1575	GCAAGAAAGCGAGCTTAGTGATACTTGTGGGCCAGGGCATTAGCCACACAGCCACCCACT	1516		
Query 5225644	TTCTGATAGGCAGCTGCACCTGGTGGGGTGAATCTTTGCAAAGTGATGGCCAGCACA	5225703		
Sbjct 1515	TTCTGATAGGCAGCTGCACCTGGTGGGGTGAATCTTTGCAAAGTGATGGCCAGCACA	1456		
Query 5225704	CAGACCAGCACGTTGCCAGGAGCTGTGGGAGGAAGATAAGAGGTATGAACATGATTAGC	5225763		
Sbjct 1455	CAGACCAGCACGTTGCCAGGAGCTGTGGGAGGAAGATAAGAGGTATGAACATGATTAGC	1396		
Query 5225764	AAAAGGGCCTAGCTTGGACTCAGAATAATCCAGCCTTATCCCAACCATAAAAATAAAGCA	5225823		
Sbjct 1395	AAAAGGGCCTAGCTTGGACTCAGAATAATCCAGCCTTATCCCAACCATAAAAATAAAGCA	1336		
Query 5225824	GAATGGTAGCTGGATTGTAGCTGCTATTAGCAATATGAAACCTCTTACATCAGTTACAAT	5225883		
Sbjct 1335	GAATGGTAGCTGGATTGTAGCTGCTATTAGCAATATGAAACCTCTTACATCAGTTACAAT	1276		
Query 5225884	TTATATGCAGAAATATTTATATGCAGAGATATTGCTATTGCCTTAACCCAGAAATATCA	5225943		
Sbjct 1275	TTATATGCAGAAATATTTATATGCAGAGATATTGCTATTGCCTTAACCCAGAAATATCA	1216		

Figure 3: Details of the first hit in BLASTN sequence alignment

The transcript ID of this sequence was accessed to identify the coordinates of the coding regions of the gene. This gene is protein-coding as is evident by the coordinates of the coding regions, which are 51-494 as seen in Figure 4.

```

gene      1..628
          /gene="HBB"
          /gene_synonym="beta-globin; CD113t-C; ECYT6"
          /note="hemoglobin subunit beta"
          /db_xref="GeneID:3043"
          /db_xref="HGNC:HGNC:4827"
          /db_xref="MIM:141900"
exon      1..142
          /gene="HBB"
          /gene_synonym="beta-globin; CD113t-C; ECYT6"
          /inference="alignment:Splign:2.1.0"
misc_feature 30..32
          /gene="HBB"
          /gene_synonym="beta-globin; CD113t-C; ECYT6"
          /note="upstream in-frame stop codon"
CDS       51..494
          /gene="HBB"
          /gene_synonym="beta-globin; CD113t-C; ECYT6"
          /note="hemoglobin, beta; beta globin chain; hemoglobin
          beta subunit"
          /codon_start=1
          /product="hemoglobin subunit beta"
          /protein_id="NP_000509.1"
          /db_xref="CCDS:CCDS7753.1"
          /db_xref="GeneID:3043"
          /db_xref="HGNC:HGNC:4827"
          /db_xref="MIM:141900"
          /translation="MVHLTPEEKSAVTALWGKVIVDEVGGEALGRLLVVYPWTQRFFESFGDLS
          TPDVAVMGNPKVKAHGKKVLAAGFSDGLAHLNLRKGTATLSELHCDKLVHVDPE
          NFRLLGNLVLCVLAHFFGKEFPPVQAAVQKVVAGVANALAHKYH"

```

Figure 4: Coding region of the gene

Based on Figure 5, this gene has three exons whose coordinates are 1-142, 143-365, and 366-628.

```

exon      1..142
          /gene="HBB"
          /gene_synonym="beta-globin; CD113t-C; ECYT6"
          /inference="alignment:Splign:2.1.0"

exon      143..365
          /gene="HBB"
          /gene_synonym="beta-globin; CD113t-C; ECYT6"
          /inference="alignment:Splign:2.1.0"

exon      366..628
          /gene="HBB"
          /gene_synonym="beta-globin; CD113t-C; ECYT6"
          /inference="alignment:Splign:2.1.0"

```

Figure 5: Coordinates of exon regions of the gene

Using the given gene ID, the genomic data of the gene was identified. This gene is present on chromosome 11 and its location is p15.4 as seen in Figure 6. It can also be seen that the name of this gene is hemoglobin subunit beta (HBB) and it is present in *Homo sapiens*.

The image shows a web browser window displaying a gene summary for HBB. The window has a title bar with a maximize icon, the text 'Summary', and a help icon. The main content is organized into several sections:

- Official Symbol:** HBB provided by HGNC
- Official Full Name:** hemoglobin subunit beta provided by HGNC
- Primary source:** HGNC:HGNC:4827
- See related:** [Ensembl:ENSG00000244734](#) [MIM:141900](#)
- Gene type:** protein coding
- RefSeq status:** REVIEWED
- Organism:** [Homo sapiens](#)
- Lineage:** Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini; Catarrhini; Hominidae; Homo
- Also known as:** ECT6; CD113t-C; beta-globin
- Summary:** The alpha (HBA) and beta (HBB) loci determine the structure of the 2 types of polypeptide chains in adult hemoglobin, Hb A. The normal adult hemoglobin tetramer consists of two alpha chains and two beta chains. Mutant beta globin causes sickle cell anemia. Absence of beta chain causes beta-zero-thalassemia. Reduced amounts of detectable beta globin causes beta-plus-thalassemia. The order of the genes in the beta-globin cluster is 5'-epsilon -- gamma-G -- gamma-A -- delta -- beta--3'. [provided by RefSeq, Jul 2008]
- Orthologs:** [all](#)

Below the summary is a section titled 'Genomic context' with a maximize icon and a help icon. It contains the text 'Location: 11p15.4' which is enclosed in a red rectangular box. To the right of this text is a link: 'See HBB in [Genome Data Viewer](#)'. At the bottom left of this section, it says 'Exon count: 3'.

Figure 6: Chromosomal location of the gene, gene name, and organism

Therefore, using the gene ID provided, the NCBI database was used to identify the associated gene and its sequence, and run a sequence alignment to identify homologies. The top 3 hits gave a 100% similarity with the HBB gene, thereby conclusively proving that the given gene ID belongs to the HBB gene.