



Identification, Analysis and Comparison of Human Alpha Tubulin Gene

Hira Mubeen¹, Muhammad Waseem Shoaib² and Shahid Raza^{1*}

¹Department of Biological Sciences, University of South Asia, Lahore, Pakistan.

²District Head Quarter (DHQ) Hospital, Faisalabad, Pakistan.

Authors' contributions

This work was carried out in collaboration between all authors. All authors read and approved the final manuscript.

Article Information

DOI: 10.9734/JABB/2016/26691

Editor(s):

- (1) Rafael A. Canas, Department of Molecular Biology and Biochemistry, Málaga University, Spain.
- (2) Anil Kumar, Professor & Head (Chair), School of Biotechnology, Devi Ahilya University, Madhya Pradesh, India.
- (3) Joana Chiang, Department of Medical Laboratory Science and Biotechnology, China Medical University, Taiwan.

Reviewers:

- (1) Bulang Gao, Hebei Medical University, PR China.
 - (2) Anonymous, Université Joseph Fourier Grenoble 1, France.
 - (3) Triveni Krishnan, National Institute of Cholera and Enteric Diseases, Kolkata, India.
- Complete Peer review History: <http://www.sciencedomain.org/review-history/15682>

Original Research Article

Received 28th April 2016
Accepted 23rd July 2016
Published 6th August 2016

ABSTRACT

Tubulins are composed of filaments or thread like structures called as microtubules which plays an important role in structural composition of cytoskeleton. Tubulin comprise of various types and subtypes. They belong to family of globular proteins. α -tubulin's are considered as a multigene family in various species. Sequence analysis of human, mouse and rat α -tubulin has enabled an updated nomenclature to be studied. Analysis of orthologues helps to study evolutionary relationships and to identify the sequence homology among all genetic variants. Researchers can predict various structural elements and can overcome many gene mapping problems.

Purpose: To study and compare the evolutionary relationships, homologs, orthologues and putative protein domains.

Materials and Methods: Sequence analysis of human, mouse, and rat α -tubulin gene has enabled an updated nomenclature to be studied. Analysis of orthologues helps to study evolutionary relationships and to identify the sequence homology among all genetic variants.

Results: Comparative study was performed by using number of different bioinformatics tools and software's like BLAST, Ensemble and Interpro. The nucleotide sequence of 1353 bp was retrieved from Genbank NCBI. The sequence was rich with many coding regions. The mRNA transcript was

*Corresponding author: E-mail: dr.raza03@gmail.com;

analyzed with 3 different splice variants. Various protein domain were analyzed which belongs to superfamily of TUBA3C gene associated protein.

Conclusion: The analysis of TUBA3C gene and protein domain helped to predict the 3D structure of related proteins. The homology and similarity studies can help to predict the evolutionary relationships.

Keywords: Tubulin gene; structure homology; orthologues; gene mapping.

1. INTRODUCTION

Microtubules are cylindrical tubes of 20-25 nm in diameter which performs various diverse functions in eukaryotic cytoskeleton. They are composed of some protofilaments and are considered as heterodimer of alpha and beta tubulin. Tubulin is found to be major constituent of microtubules. Hence, the genes which code for these microtubules are known to be part of the tubulin superfamily. It binds two moles of GTP one at an exchangeable site on the beta chain and other at a non-exchangeable site on the alpha chain. There are six distinct families of tubulin. About six human genes encoding different tubulin subunits are observed from previous studies [1]. Similarly, genes from the alpha, beta and gamma tubulin families are found to be rich in mostly all eukaryotes. Alpha and beta tubulins are more commonly found in microtubules. However, gamma tubulin is mainly involved in the nucleation of microtubule assembly. According to some previous studies, multiple alpha and beta tubulin genes are found to be highly conserved among and between species. A number of studies suggested, most epithelial tumor cells which are highly expressed as isoform of tubulin is beta tubulin, which is encoded by the TUBB gene [2-4].

Multiple mutations associated with alpha tubulin gene have been reported. Some of the α -tubulin mutations have found to be associated with lissencephaly and impaired neuronal migration [5]. Moreover, γ -tubulin is necessary for nucleation of the microtubule assembly. On the other side, the role of the remaining tubulin families are not well characterized yet [6,7]. Human genome consists of variety of the α -tubulin genes. The degree of tissue specificity also varies in expression of some α -tubulins, with several different mouse and human α -tubulins [8,9].

To study about various orthologues of TUBA3C, whole genome can be analyzed to know about all the predicted alpha tubulin gene families. The studies through HGNC and MGNC nomenclature data will help to study the evolutionary

relationships, sequence comparisons, predicted homologs, orthologues and mainly about protein domains [10].

2. TUBA3C GENE

TUBA3C (Tubulin, Alpha 3c) is found as a Protein Coding gene. TUBA 3C mainly encodes a protein 99% identical to the mouse testis-specific Tuba3 and Tuba7 gene products. Multiple diseases are found to be related with this gene like ectodermal dysplasia 2 and dfnb1. Genome annotations related to this selected gene include GTP binding and structural constituent of cytoskeleton. The study was performed to analyze and compare the TUBA3C gene with its various orthologues at gene and protein level. The conserved domains of protein were also analyzed by using various bioinformatics tools.

3. MATERIALS AND METHODS

Tubulins are considered as major building blocks of microtubules. They have been divided into several subunits mainly alpha and beta subunits. These subunits play a very important role in assembly of various microtubules. Sequence comparison of tubulin gene with other predicted orthologues was performed by using various bioinformatics tools and software's. The methodology used to retrieve, compare and analyze the TUBA3C gene and its associated protein is discussed below.

3.1 Sequence Retrieval of TUBA3C Gene

Sequence of TUBA3C gene was retrieved from NCBI-Genbank data repository. The sequence features are studied. All coding regions are highlighted. Further, this sequence was translated into protein.

3.2 Chromosomal Map Analysis

Chromosomal Map of the TUBA3C gene was analyzed by using Ensemble software. The complete location of gene was indicated on chromosome.

3.3 Identification of Orthologues

Orthologues for TUBA3C gene were identified by using Ensemble software. A number of different orthologues were found from mouse, rats, horse, chimpanzee etc.

3.4 Protein Domain Analysis through InterPro Tool

InterPro was used to study various protein matches and signatures. The putative conserved domain of TUBA3C gene was found by using InterPro data repository.

4. RESULTS

Microtubules are known to be functionally distinct structures involved in cellular processes like mitosis, cell movement and cell division. The important components of microtubules includes different isoforms of alpha and beta tubulins. The translated sequence of selected gene was obtained to analyze the conserved regions. Five were identified in the selected gene. Gene transcription produces 4 different mRNAs with 3 alternatively spliced variants and 1 unspliced form. The mRNAs appear to differ by truncation of the 5' end, overlapping exons with different boundaries. The 3 spliced and the unspliced mRNAs putatively encode good proteins, altogether 4 different isoforms (2 complete, 1 COOH complete, 1 partial), some containing domains Tubulin family, GTPase domain, Tubulin C-terminal domain [Pfam].

4.1 Sequence Retrieval from NCBI

Sequence analysis of TUBA3C gene was performed with bioinformatics analysis tools. Coding regions were analyzed. The nucleotide sequence of 1352 bp was obtained from Genbank NCBI. The highlighted region shows the coding regions of gene.

4.2 Chromosome Map

The chromosome map was obtained from ensemble software. The TUBA3C gene was located on chromosome number 13. The Fig. 3 shows the position of gene on chromosome.

4.3 Identification of Orthologues

The TUBA3C gene was analyzed with other gene variants in different species. All vary in size and are present on different chromosomes. Total 64 different orthologues have similarity with TUBA3C gene. Few of them are given below in the Table 1.

4.4 Consensus CDS Protein Set

The TUBA3C nucleotide sequence was assigned ENST00000400113 ID by ensemble. Similarly, the ID for its associated protein is ENS00000382982. These gene coding proteins are submitted in specific consensus protein database.

```
>ATGCGTGAGTGTATCTCTATCCACGTGGGGCAGGCAGGAGTCCAGATCGGCAATGCCTGCTGGGAAC
TGTACTGCCTGGAACATGGAATTCAGCCCGATGGTCAGATGCCAAGTGATAAAACCATTGGTGGTGGGG
ACGACTCCTTCAACACGTTCTTCAGTGAGACTGGAGCTGGCAAGCACGTGCCAGAGCAGTTTGTGGAC
CTGGAGCCCACTGTGGTTCGATGAAGTGCACAGGAACCTATAGGCAGCTCTCCACCAGAGCAGCTG
ATCACCAGGGAAGGAAGATGCGGCCAATAATTACGCCAGAGGCCATTACACCATCGGAGATCGTCGACCT
GGTCTGACCGGATCCGCAAACTGGCGGATCTGTGCACGGGACTGCAGGCTCTCATCTCCACAGTTT
TGGGGGTGGCACTGGCTCTGGGTTCCGATCTGTGTCATGGAGCGGCTCAGTGGATTACGGCAAGAAG
TCCAAGCTAGAATTTGCCATTTACCCAGCCCCCAGGTCCTCCACGGCCGTGGTGGAGCCCTACAACCTC
ATCCTGACCACCCACAGACCCTGGAACATTCTGACTGTGCCTTCATGGTCGACAATGAAGCCATCTATG
ACATATGTCGGCGCAACCTGGACATCGAGCGTCCCACGTACACCAACCTCAATCGCCTGATTGGGCAGA
TCGTGTCTCCATCACGGCCTCCCTGCGATTGACGGGGCCCTGAATGTGGACTTGACGGAATTCCAGA
CAAACCTAGTGCCGTACCCCGCATCCACTTCCCCTGGCCACCTACGCCCCGGTCATCTCAGCCGAG
AAGGCCACACGAGCAGCTGTCCGTGGCTGAGATCACCAATGCCTGCTTCGAGCCAGCCAATCAGAG
GTCAAGTGTGACCCTGCCACGGCAAGTACATGGCCTGCTGCATGTTGTACAGGGGGGATGTGGTCCC
GAAAGATGTCAACGCGGCCATCGCCACCATCAAGACCAAGCGCACCATCCAGTTTGTAGATTGGTGCCC
AACTGGATTTAAGGTGGGCATTAACCTACCAGCCCCCAGGTTGGTCCCTGGGGGAGACCTGGCCAAGG
TGCAGCGGGCTGTGTGCATGCTGAGCAACACCACGGCCATCGCGGAGGCCTGGGCTCGCCTGGACCA
TAAATTCTCTGAGGCCCGCGAGGACCTGGCAGCTGAGAGGATTATGAAGAGGTGGGCGTGGAT
TCCGTGGAAGCCGAGGCTGAAGAAGGTGAAGAATACTGA
```

Fig. 1. Shows the nucleotide sequence of TUBA3C gene obtained from GenBank-NCBI. The highlighted region shows the coding regions

4.5 Protein Domain

Proteins are macromolecules made up of some functional folding units referred as domains. The TUBA3C protein was searched in Interpro protein database for identification of various protein

domains and signature with assigned accession ID i.e, Q13748. The length of protein sequence was 450 aa. All domains are part of different protein families with several conserved patterns and signatures. Some of the domains are shown in Fig. 5.

Translated Protein Sequence

```
MRECISIHVGQAGVQIGNACWELYCLEHGIQPDGQMPSDKTIGGGDDSFNTFFSETGAGKHVPRAVFVDL
EPTVWDEVRTGTYRQLFHPEQLITGKEDAANNYARGHYTIGKEIVDLVLDRIKRLADLCTGLQGFLIFHS
FGGGTGSGFASLLMERLSVDYGGKSKLEFAIYPAPQVSTAVVEPYNSILTHTTLEHSDCAFMDNEAIY
DICRRNLDIERPTYTNLRLIGQIVSSITASLRFDGALNVDLTFQTNLVPYPRIHFLATYAPVISAEK
AYHEQLSVAEITNACFEPANQMVKCDPRHGKYMCCMLYRGDVPKDVNAAIATIKTKRTIQFVDWCPTG
FKVGINYQPPTVPPGGDLAKVQRAVCMLSNTTATAEAWARLDHKFDLMYAKRAVHVHWYVGEEMEEGFSE
AREDLAALAEKDYEEVGVDSVEAEAEEGEEY
```

Fig. 2. Shows the translated protein sequence of TUBA3C gene obtained from GenBank-NCBI

Chromosome 13: 19,173,770-19,181,852

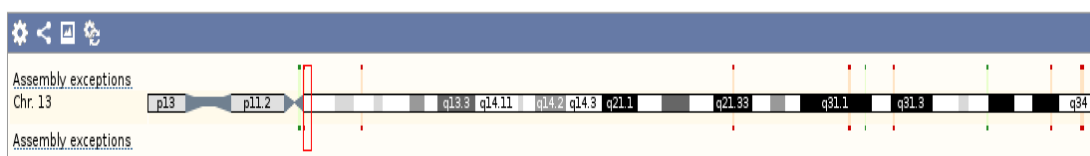


Fig. 3. Shows the chromosome map of TUBA3C from ensemble software

Sequence IDs included in CCDS 9284.1

Original	Current	Source	Nucleotide ID	Protein ID	Status in CCDS	Seq. Status
✓	✓	EBI,WTSI	ENST00000400113	ENSP00000382982	Accepted	alive
✓	✓	EBI,WTSI	OTTHUMT00000044007	OTTHUMP00000018071	Accepted	alive
✓	✓	NCBI	NM_006001.2	NP_005992.1	Accepted	alive

RefSeq	Length	Related UniProtKB/SwissProt	Length	Identity	Gaps	Mismatches
NP_005992.1	450	Q13748-1	450	100%	0	0

Fig. 4. Shows the nucleotide and protein sequence of TUBA3C gene submitted in consensus protein database

Table 1. Shows the reported orthologues of TUBA3C gene

Sr. No.	Gene ID	Gene name	Species	Location
1.	7278	TUBA3C	<i>Homo sapiens</i>	Chromosome 13
2.	707215	TUBA3C	<i>Rhesus monkey</i>	Chromosome 17
3.	107209044	TUBA3C	<i>Parus major</i>	Chromosome 9
4.	106826191	TUBA3C	<i>Equus asinus</i>	-----
5.	104356476	TUBA3	<i>Tyto alba</i>	-----
7.	101062302	tuba3c	<i>Takifugurubripes</i>	Chromosome 2
8.	100993591	TUBA3C	<i>Pan pansicus</i>	Chromosome 13
9.	100585215	TUBA3C	<i>Nomascus</i>	Chromosome 5
10.	100434822	TUBA3C	<i>Pongo abelii</i>	Chromosome 13
11.	465169	TUBA3C	<i>Pan troglodytes</i>	Chromosome 13

Tubulin alpha-3C/D chain (Q13748)

Accession [Q13748](#) (TBA3C_HUMAN)
Species Homo sapiens (Human)
Length 450 amino acids (complete)

Protein family membership

Tubulin (IPR000217)
Alpha tubulin (IPR002452)

Domains and repeats

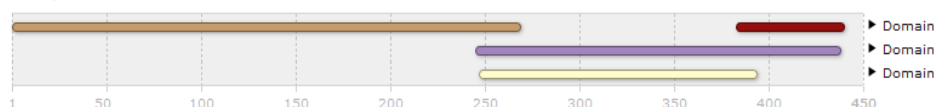


Fig. 5. Shows the putative protein domains of TUBA3C human gene with tubulin protein obtained from interpro protein databank

5. DISCUSSION AND CONCLUSION

Tubulin belongs to family of globular proteins superfamily. It has multiple types mainly two important, α and β -tubulins which tend to polymerize into microtubules. These are major components of cytoskeleton and function in many essential cellular processes, including mitosis. Many post translational modifications occur at C terminal. Some glutamate residues at the C-terminus are known to be polyglutamylated. Another modification, glycylation is mainly limited to tubulin incorporated into cilia and flagella. The precise function of such modifications is still unclear but they play an important role to regulate the assembly of microtubules. An alpha-tubulin gene was identified and cloned as the human homolog, which showed identity to an EST named TUBA2. The TUBA2 gene encodes a protein which consists of 450-amino acid that is identical to some mouse alpha-tubulins [11]. Moreover, it was concluded that the TUBA2 gene contains 5 exons and spans at least 6.5 kb. All 4 of its introns are located at conserved positions. The 3-prime untranslated region of TUBA2 is 85% homologous to that of TUBA1. The analysis of TUBA3C gene and protein domain helped to predict the 3D structure of selected protein and its related proteins. Finally proteins from this gene may be modulated by acetylation, mono-methylation; phosphorylation and ubiquitination.

According to some studies, the human tubulin gene family consists of 15 to 20 dispersed genes

including pseudogenes [12]. It was observed that position of introns in humans and other organisms especially rats are quite identical. However, genes can be differentiated on the basis of 3-prime untranslated regions (UTRs). Alpha tubulin genes can be classified according to similarity of some C-terminal motifs. The analysis of selected gene shows there are 5 exons and many introns which are located at some conserved positions. However, the similarity searches indicates that the human and mouse tubulin genes are 88% similar [13].

The analysis of TUBA3C gene showed its genetic variants in different organisms. All are located on different chromosomes and different positions. Each gene was having some similar characteristics according to their biological emergence and evolution. The evolutionary relationships and lineages provides detailed information about related ancestors. The TUBA3C protein was identified with different putative conserved domains. All domains were analyzed by using Interpro protein network. The first conserved domain was found at 270aa position. Similarly, the second domain was found from 280aa to 450aa position. All predicted protein domains showed different signatures and patterns. The crystal structure of selected protein helps to study the 3D configuration of tubulin protein. All TUBA3C genes have different configuration patterns. Therefore, the analysis of TUBA3C protein will lead to a thorough understanding of the tubulin gene, its evolutionary pathways, genetic location,

occurrence, its similarity and homology with other tubulin gene variants, identification of orthologues and most importantly the conserved functional domains within the protein region.

COMPETING INTERESTS

Authors have declared that no competing interests exist.

REFERENCES

1. Luduena RF. Multiple forms of tubulin: Different gene products and covalent modifications. *Int Rev Cytol.* 1998;178: 207–75.
2. Lee MG, Lewis SA, Wilde CD, Cowan NJ. Evolutionary history of a multigene family: An expressed human beta-tubulin gene and three processed pseudogenes. *Cell.* 1983;33:477–87.
3. Gonzalez-Garay ML, Chang L, Blade K, Menick DR, Cabral F. A beta-tubulin leucine cluster involved in microtubule assembly and paclitaxel resistance. *J Biol Chem.* 1999;274:23875–82.
4. Giannakakou P, Sackett DL, Kang YK, Zhan Z, Buters JT, Fojo T, et al. Paclitaxel-resistant human ovarian cancer cells have mutant beta-tubulins that exhibit impaired paclitaxel-driven polymerization. *J Biol Chem.* 1997;272:17118–25.
5. Keays DA, et al. Mutations in alpha-tubulin cause abnormal neuronal migration in mice and lissencephaly in humans, *cell* 128. 2007;45–57.
6. Kierszenbaum AL. Sperm axoneme: A tale of tubulin posttranslation diversity, *Mol. Reprod. Dev.* 2002;62:1–3.
7. McKean PG, Vaughan S, Gull K. The extended tubulin superfamily. *J. Cell Sci.* 2001;114:2723-2733.
8. Villasante A, et al. Six mouse alpha-tubulin mRNAs encode five distinct isoforms: Testis-specific expression of two sister genes. *Mol. Cell. Biol.* 1986;6:2409–2419.
9. Hall JL, Cowan NJ. Structural features and restricted expression of a human alpha-tubulin gene. *Nucleic Acids Res.* 1985;13:207-223.
10. Wain HM, et al. Guidelines for human gene nomenclature. *Genomics.* 2002;79: 464–470.
11. Dode C, Weil D, Levilliers J, Crozet F, Chaib H, Levi-Acobas F, Guilford P, Petit C. Sequence characterization of a newly identified human alpha-tubulin gene (TUBA2). *Genomics.* 1998;47:125-130.
12. Lewis SA, Cowan NJ. Tubulin genes: Structure, expression, and regulation. In: Avila J, (ed.). *Microtubule proteins.* Boca Raton: CRC Press, Inc. 1990;37-66.
13. Dode C, Weil D, Levilliers J, Crozet F, Chaib H, Levi-Acobas F, Guilford P, Petit C. Sequence characterization of a newly identified human alpha-tubulin gene (TUBA2). *Genomics.* 1998;47:125-130.

© 2016 Mubeen et al.; This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Peer-review history:

The peer review history for this paper can be accessed here:
<http://sciencedomain.org/review-history/15682>