Atlas of Genetics and Cytogenetics in Oncology and Haematology

OPEN ACCESS JOURNAL AT INIST-CNRS

Gene Section

Short Communication

C2orf3

Masato Takimoto, Mao Peizhong

Institute for Genetic Medicine, Hokkaido University, Sapporo, Japan; takimoto@igm.hokudai.ac.jp (MT); Dept of Cell Developmental and Cancer Biology, School of Medicine and Division of Neuroscience, Oregon National Primate Research Center, Oregon Health Science University, Oregon, USA maop@ohsu.edu (MP)

Published in Atlas Database: May 2017

Online updated version : http://AtlasGeneticsOncology.org/Genes/C2orf3ID54327ch2p12.html
Printable original version : http://documents.irevues.inist.fr/bitstream/handle/2042/68875/05-2017-C2orf3ID54327ch2p12.pdf
DOI: 10.4267/2042/68875

Abstract

Review on C2orf3, with data on DNA, on the protein encoded, and where the gene isimplicated.

Keywords

post-splicing, turnover of mRNA, lariat intron, dyslexia

Other names

GCFC2 GC-rich sequence DNA-binding factor 2

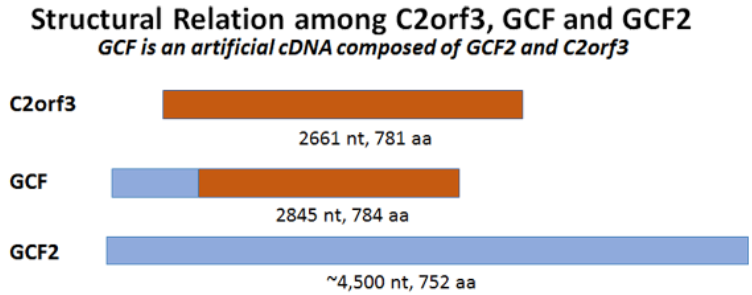
HGNC (Hugo)

GCFC2

Location

2p12

Identity



Structural Relation among C2orf3, GCF and GCF2. GCF is an artificial cDNA composed of GCF2 and C2orf3

Note

C2orf3 gene was discovered as a C-terminal side of GCF cDNA, which is an artificial chimeric one (Kageyama and Pastan, 1989;Takimoto et al., 1999) . As C2orf3 protein is a factor that plays a role in splicing of mRNA (Yoshimoto et al., 2014) and the protein encoded by a N-terminal side of GCF cDNA, which is a bona fide transcriptional repressor termed GCF2, binds to the GC rich element of DNA (Reed et al., 1998), the nomenclature of this gene as GCFC2 is inappropriate.

Major data bases, such as provided by University of California Santa Cruz (UCSC), National Center for Biotechnology Information (NCBI) and Hugo Gene Nomenclature Committee (HGNC), use GCFC2 to describe this gene. For example, UCSC genome browser describes this gene as GCFC2, which is

present next to MRP19 gene in the chromosomal 2p12 region (Anthoni et al., 2007).
As C2orf3 does not bind to the GC-rich sequence (Takimoto et al., 1999), the use of the term GCFC2 for this gene is misleading and we think it should not be used.

DNA/RNA

Description

Genomic DNA
Genomic size of DNA is about 50 kbp and consists of 17 exons. >

cDNA
GCF was originally discovered as a gene that encodes a transcriptional repressor that binds to a GC-rich sequence of the regulatory regions of human EGF-R gene (Kageyama and Pastan, 1989). However, a following study showed that GCF cDNA is an artificial fusion molecule between two different cDNAs; The 5' side of the cDNA encodes a highly basic region with the sequence-specific binding activity to the GC-rich region and the 3' side encodes most of C2orf3 protein (Takimoto et al., 1999). Reed et al. cloned a full length of the former cDNA that has transcriptional repressor activity with the sequence-specific DNA binding ability, and named this cDNA as GCF2. Following studies had confirmed the transcriptional repressive activity of GCF2 (Shibutani M et al., 1998: Khachigian LM et al., 1999). The full length of C2orf3 cDNA was cloned and its sequence was determined. The cDNA is composed of 2661 nucleotides and encodes a protein of 781 amino acids. The encoded protein does not contain the highly basic region of the published GCF and has no sequence-specific DNA binding activity to the GC-rich sequence (Takimoto et al., 1999)

cDNA sequence: (GenBank: AB026911)
GGGCGGGCACTGAAGCTGCGGCTTGCGGTT
CAGCGGGTTCTAGGGCGCCGGGCGCTCGGG
CCTCGGCCATGGCTCACAGGCCGAAAAGGA
CTTTTCGGCA
GCGCGCGGCTGATTCCAGCGACAGCGATGG
CGCCGAGGAGTCGCCTGCTGAGCCTGGGGC
GCCGAGGGAACTTCCGGTCCCGGGTTCTGC
GGAGGAAGAG
CCGCCCTCTGGAGGAGGCCGCGCGCAGGTG
GCGGGACTGCCCCACCGGGTTCGGGGCCCT
CGTGGCCGGGGCCGGGTCTGGGCGAGCTCC
CGGCGTGCCA
CCAAAGCGGCTCCCCGCGCGGACGAAGGCT
CAGAATCCAGAACCCTTGATGTGTCCACAG
ATGAAGAGGATAAAATACATCACTCCTCAG
AAAGTAAGGA
TGATCAGGGTTTGTCTTCTGACAGTTCTAGC
TCTCTTGGAGAAAAAGAACTTTCATCAACA
GTTAAGATCCCAGATGCAGCTTTTATTCAG
GCAGCCCGC
AGAAAACGTGAATTGGCCAGGGCCCAAGAT
GACTATATTTCTTTGGATGTACAACATACCT
CCTCCATCTCTGGTATGAAGAGAGAGAGCG
AAGATGACC
CTGAGAGTGAGCCTGATGACCATGAAAAGA
GAATACCATTTACTCTAAGACCTCAAACAC
TTAGACAAAGGATGGCTGAGGAATCAATAA
GCAGAAATGA
AGAAACAAGTGAAGAAAGTCAGGAAGATG
AAAAGCAAGATACTTGGGAACAACAGCAA
ATGAGGAAAGCAGTTAAAATCATAGAGGA
AAGAGACATAGAT
CTTTCCTGTGGCAGTGGATCTTCAAAAGTG
AAGAAATTTGATACTTCCATTTCATTTCCGC
CAGTAAATTTAGAAATTATAAAGAAGCAAT
TAAATACTA
GATTAACATTACTACAGGAAACTCACCGCT
CACACCTGAGGGAGTATGAAAAATACGTAC
AAGATGTCAAAAGCTCAAAGAGTACCATCC
AGAACCTAGA
GAGTTCATCAAATCAAGCTCTAAATTGTAA
ATTCTATAAAAGCATGAAAATTTATGTGGA
AAATTTAATTGACTGCCTTAATGAAAAGAT
TATCAACATC
CAAGAAATAGAATCATCCATGCATGCACTC
CTTTTAAAACAAGCTATGACCTTTATGAAA
CGCAGGCAAGATGAATTAAAACATGAATCA
ACGTATTTAC
AACAGTTATCACGCAAAGATGAGACATCCA
CAAGTGGAAACTTCTCAGTAGATGAAAAAA
CTCAGTGGATTTTAGAAGAGATTGAATCTC
GAAGGACAAA
AAGAAGACAAGCAAGGGTGCTTTCTGGGAA
TTGTAACCATCAGGAAGGAACATCTAGTGA
TGATGAACTGCCTTCAGCAGAGATGATTGA
CTTCCAAAAA
AGCCAAGGTGACATTTTACAGAAACAGAAG
AAAGTTTTTGAAGAAGTGCAAGATGATTTT
TGTAACATCCAGAATATTTTGTTGAAATTTC
AGCAATGGC
GAGAAAAGTTTCCTGACTCCTATTATGAAG
CTTTCATTAGTTTATGCATACCAAAGCTTTT
AAATCCCCTAATACGAGTTCAGTTGATTGA
TTGGAATCC
TCTTAAGTTGGAATCCACAGGTTTAAAAGA
GATGCCATGGTTCAAATCTGTAGAAGAATT
TATGGATAGCAGTGTAGAAGATTCAAAGAA
GGAAAGTAGT
TCAGATAAAAAAGTCTTGTCTGCAATCATC
AACAAAACAATTATTCCCCGACTTACAGAC
TTTGTAGAATTCCTTTGGGATCCTTTGTCAA
CCTCACAGA
CAACAAGTTTAATAACACATTGCAGAGTGA
TTCTTGAAGAACATTCCACTTGTGAAAATG
AAGTTAGTAAAAGCAGACAGGATTTACTTA
AATCCATTGT
TTCAAGAATGAAAAAGGCAGTAGAAGATG
ATGTTTTTATTCCTCTGTATCCAAAGAGTGC
TGTAGAAAACAAAACATCACCTCATTCAAA

GTTCCAAGAA
AGACAGTTCTGGTCAGGCCTAAAGCTCTTC
CGCAATATTCTTCTTTGGAATGGACTCCTTA
CAGATGACACCTTGCAAGAACTAGGACTAG
GGAAGCTGC
TAAATCGTTACCTTATTATAGCACTTCTCAA
TGCCACACCTGGGCCAGATGTGGTTAAAAA
GTGCAACCAGGTAGCAGCATGTCTACCAGA
AAAATGGTT
TGAAAATTCTGCCATGAGGACATCTATTCC
ACAGCTAGAAAACTTCATTCAGTTTTTATTG
CAGTCTGCACATAAATTATCTAGAAGTGAA
TTCAGGGAT
GAAGTCGAAGAAATAATTCTTATTTTGGTG
AAAATAAAAGCTTTGAATCAAGCAGAATCC
TTCATAGGAGAGCATCACCTAGACCATCTT
AAATCACTAA
TTAAAGAAGATTGAATAAACTTTATTGGAA
AATGCTAAAATTTTAATATAGTTACACTCA
GTTCCTTTGTTTGAGAAGAAGCTGGTGCCTC
TCTCTTCTT
TATTCCCTGTAATAGAAGGTAGGATTTGAA
AAAAAGCAGGACTCCACCTCTGTATTCCCC
CGTGCTTTACCTTCTGGCATCATGAAAAGCT
GCCATGATT
CTGTGGTGTTCTAAGGAATTAAATGCACTG
GAGCTTTAAGAGCTCAACGTGTTTCCCTTTG

Transcription

It is suggested that MRPL1 gene which is located closely to C2orf3 with head-head orientation and that they are co-regulated (Anthoni et al., 2007). There are several alternative spliced forms for mRNA expression: The initial report on a cDNA that encompass the full coding region is 2661bp in length (Takimoto et al., 1999).

Protein

Description

Encodes 781 amino acids

Amino acid sequence: (Uniprot : P16383)

MAHRPKRTFRQRAADSSDSDGAEESPAEPGA
PRELPVPGSAEEEPPSGGGRAQVAGLPHRVRG
PRGRGRVWASSRRATKAAPRADEGSESRTLD
VSTDEE
DKIHHSSESKDDQGLSSDSSSSLGEKELSSTVK
IPDAAFIQAARRKRELARAQDDYISLDVQHTS
SISGMKRESEDDPESEPDDHEKRIPFTLRPQTL
RQ
RMAEESISRNEETSEESQEDEKQDTWEQQQM
RKAVKIIEERDIDLSCGSGSSKVKKFDTSISFPP
VNLEIIKKQLNTRLTLLQETHRSHLREYEKYV
QDV
KSSKSTIQNLESSSNQALNCKFYKSMKIYVEN
LIDCLNEKIINIQEIESSMHALLLKQAMTFMKR
RQDELKHESTYLQQLSRKDETSTSGNFSVDEK
TQW
ILEEIESRRTKRRQARVLSGNCNHQEGTSSDD
ELPSAEMIDFQKSQGDILQKQKKVFEEVQDDF
CNIQNILLKFQQWREKFPDSYYEAFISLCIPKL
LNP
LIRVQLIDWNPLKLESTGLKEMPWFKSVEEF
MDSSVEDSKKESSSDKKVLSAIINKTIIPRLTD
FVEFLWDPLSTSQTTSLITHCRVILEEHSTCEN
EVS
KSRQDLLKSIVSRMKKAVEDDVFIPLYPKSAV
ENKTSPHSKFQERQFWSGLKLFRNILLWNGLL
TDDTLQELGLGKLLNRYLIIALLNATPGPDVV
KKCN
QVAACLPEKWFENSAMRTSIPQLENFIQFLLQ
SAHKLSRSEFRDEVEEIILILVKIKALNQAESFI
GEHHLDHLKSLIKED

Expression

C2orf3 protein with molecular weight of 89 kD are observed in human cancer cell line, and localizes in nucleoplasm and nucleolus.

Function

C2orf3 plays a role in pre-mRNA splicing, by forming a complex with DHX15 (hPrp43) and TFIP11 (Yoshimoto et al., 2014). As these proteins are present in post-splicing intron complex, C2orf3 protein may play a role in post-splicing turnover of mRNA. The study with and antibody specific to C2ofr3 protein showed that this protein is present nucleoplasm and nucleoli.

After splicing reaction, pre-mRNA releases intron RNA complex, which contains uridine-rich small nuclear RNAs (snRNAs; U1, U2, U4, U5 and U6) PRPF19complex, hnRNP proteins and TFIP11. A RNA helicase hPrp43 removes the several factors from the complex, leaving lariat RNA intron, which is then subject to linearization by a debranching enzyme DBR1 (Wen et al., 2008; Yoshimoto et al., 2009).C2orf3 protein was shown to form a complex with tuftelin-interacting protein (TFIP11) and hPrp43, which play a role in post-splicing turnover of mRNA. Through its amino terminal, TFIP11 binds to a RNA helicase hPrp43 that plays a role in the dissociation of snRNAs from a lariat intron in vitro. C2orf3 preferentially associates with lariat intron in the splicing reaction and C2orf3-deleted nuclear extracts showed a significant repression of splicing of pre-mRNA in vitro (Yoshimoto et al., 2014). The presence of C2orf3 protein in nucleoli suggest a potential role in rRNA processing/or nucleoli structure (Yoshimoto et al., 2014).

Implicated in

Although it is not conclusive, C2orf3 is suggested to be a causing gene for dyslexia.

Dyslexia

The locus containing the C2orf3 gene on Chromosome 2p12 has been shown to link to dyslexia. It had been reported that the genomic regions responsive for human dyslexia, such DYXC1/EKN1, KIAA0319 and DCDC, and RUBO1, are located on human chromosome 15, 6

and 3, respectively (McGrath et al., 2006). In 1999, It was shown that a new region for dyslexia, DYX3, on human chromosome 2 was identified (Fagerheim et al., 1999). Subsequently, a study on Finnish and German families disclosed that DYX3 was present on chromosome 2p12, spanning 157 kbp. It was shown that there are only three genes, FLJ1339, MRPL1 and C2ORF3, in this region and that the latter two genes are closely located with positions in a head-to-head manner respective for transcriptional orientation, suggesting that both genes are transcriptionally co-regulated (Anthoni et al., 2007). Further analyses on several affected families revealed an overlapping region with risk haplotype within the 157 kbp region, delineating to 16 kbp, which located in an intergenic region between FLJ1339 and MRPL1/C2ORF3 genes. There is no SNP marker in the coding regions of MRPL1 and C2ORF3 genes by which coding change correlated with dyslexia, and the expressions of both genes are significant lower in carriers with risk haplotype compared with non-carriers. These results suggested that the 16 kbp region plays a role for transcriptional regulatory element and mutation in this element might lead to reduced expression of the genes, which could be a cause of dyslexia. While the expression of FLJ1339 in human brain, the expressions of MRPL1/C2ORF3 are high and significantly correlated with those of other dyslexia candidate genes of whichexpressions are also high in brain. Especially, the expression of C2ORF3 was correlated across the different parts of brain with those of other dyslexia candidate genes, DYXC1, RUBO1 and DCDC2 (Anthoni et al., 2007). Neuroimaging analyses revealed a significant association between a SNP marker and white matter volume of the posterior parts of the corpus callosum and cingulum (Scerri et al., 2012). In contrary to the studies described above, the studies on the populations of Australia and Inida showed non-significant association for the SNP marker for MRPL1/C2ORF3 with dyslexia (Paracchini et al., 2011 : Venkatesh et al., 2013).

To be noted

Previouly described GCF cDNA is an artificial fusion molecule between two different cDNAs, in which the 3' side of the molecule is derived from C2orf3 cDNA.

References

Anthoni H, Zucchelli M, Matsson H, Müller-Myhsok B, Fransson I, Schumacher J, Massinen S, Onkamo P, Warnke A, Griesemann H, Hoffmann P, Nopola-Hemmi J, Lyytinen H, Schulte-Körne G, Kere J, Nöthen MM, Peyrard-Janvid M. A locus on 2p12 containing the co-regulated MRPL19 and C2ORF3 genes is associated to dyslexia. Hum Mol Genet. 2007 Mar 15;16(6):667-77

Fagerheim T, Raeymaekers P, Tønnessen FE, Pedersen M, Tranebjaerg L, Lubs HA. A new gene (DYX3) for dyslexia is located on chromosome 2. J Med Genet. 1999 Sep;36(9):664-9

Kageyama R, Pastan I. Molecular cloning and characterization of a human DNA binding factor that represses transcription. Cell. 1989 Dec 1;59(5):815-25

Khachigian LM, Santiago FS, Rafty LA, Chan OL, Delbridge GJ, Bobik A, Collins T, Johnson AC. GC factor 2 represses platelet-derived growth factor A-chain gene transcription and is itself induced by arterial injury. Circ Res. 1999 Jun 11;84(11):1258-67

McGrath LM, Smith SD, Pennington BF. Breakthroughs in the search for dyslexia candidate genes. Trends Mol Med. 2006 Jul;12(7):333-41

Paracchini S, Ang QW, Stanley FJ, Monaco AP, Pennell CE, Whitehouse AJ. Analysis of dyslexia candidate genes in the Raine cohort representing the general Australian population. Genes Brain Behav. 2011 Mar;10(2):158-65

Reed AL, Yamazaki H, Kaufman JD, Rubinstein Y, Murphy B, Johnson AC. Molecular cloning and characterization of a transcription regulator with homology to GC-binding factor. J Biol Chem. 1998 Aug 21;273(34):21594-602

Scerri TS, Darki F, Newbury DF, Whitehouse AJ, Peyrard-Janvid M, Matsson H, Ang QW, Pennell CE, Ring S, Stein J, Morris AP, Monaco AP, Kere J, Talcott JB, Klingberg T, Paracchini S. The dyslexia candidate locus on 2p12 is associated with general cognitive ability and white matter structure. PLoS One. 2012;7(11):e50321

Shibutani M, Lazarovici P, Johnson AC, Katagiri Y, Guroff G. Transcriptional down-regulation of epidermal growth factor receptors by nerve growth factor treatment of PC12 cells. J Biol Chem. 1998 Mar 20;273(12):6878-84

Takimoto M, Mao P, Wei G, Yamazaki H, Miura T, Johnson AC, Kuzumaki N. Molecular analysis of the GCF gene identifies revisions to the cDNA and amino acid sequences(1) Biochim Biophys Acta 1999 Oct 6;1447(1):125-31

Venkatesh SK, Siddaiah A, Padakannaya P, Ramachandra NB. Lack of association between genetic polymorphisms in ROBO1, MRPL19/C2ORF3 and THEM2 with developmental dyslexia Gene 2013 Oct 25;529(2):215-9

Wen X, Tannukit S, Paine ML. TFIP11 interacts with mDEAH9, an RNA helicase involved in spliceosome disassembly Int J Mol Sci 2008 Nov;9(11):2105-13

Yoshimoto R, Okawa K, Yoshida M, Ohno M, Kataoka N. Identification of a novel component C2ORF3 in the lariat-intron complex: lack of C2ORF3 interferes with pre-mRNA splicing via intron turnover pathway Genes Cells 2014 Jan;19(1):78-87

This article should be referenced as such:

Takimoto M, Peizhong M. C2orf3. : Atlas Genet Cytogenet Oncol Haematol. 2018; 22(3): 83-86.