Cloning, chromosome localization and features of a novel human gene, *MATH2*

LINGCHEN GUO[1], MIN JIANG[1], YUSHU MA[1], HAIPENG CHENG[1], XIAOHUA NI[1], YANGSHENG JIN[2], YI XIE[1]* and YUMIN MAO[1]*

[1]*State Key Laboratory of Genetic Engineering, Institute of Genetics, School of Life Sciences, Fudan University, Shanghai 200433, People's Republic of China*
[2]*United Gene Holdings, Ltd, Shanghai 200092, People's Republic of China*

Abstract

We report cloning and some features of a novel human gene, *MATH2*, which encodes a protein of 337 amino acid residues with a basic helix–loop–helix domain and exhibits 98% similarity to mouse *Math2*. Results of Northern blot analysis revealed two transcripts of the *MATH2* gene of 1.7 kb and 2.4 kb in human brain. We localized *MATH2* to chromosome 7 at 7p14–15 by matching with the Human Genome Sequence Database. Human *MATH2* and mouse *Math2* may have the same functions in the nervous system.

[Guo L., Jiang M., Ma Y., Cheng H., Ni X., Jin Y., Xie Y. and Mao Y. 2002 Cloning, chromosome localization and features of a novel human gene, *MATH2*. *J. Genet.* **81**, 13–17]

Introduction

Transcription factors are DNA-binding proteins that regulate transcription of genes. Some transcription factors contain a basic helix–loop–helix (bHLH) domain. Examples of members of the bHLH protein family are Mash1 (Lo *et al*. 1991) and Hes1 (Sasai *et al*. 1992). Hes1, synthesized at high levels in undifferentiated neural precursor cells, is a mammalian homologue of the *Drosophila* pair-rule protein Hairy. Persistent expression of *Hes1* inhibits neural differentiation in the central nervous system (Ishibashi *et al*. 1994). Mash1 is a rat homologue of the *Drosophila* achaete-scute complex (AS-C) products. As an important regulator of neurogenesis in the ventral telencephalon, Mash1 is specifically synthesized in neural precursors (Casarosa *et al*. 1999). The fact that mutants in *Mash1* exhibit no apparent abnormalities in the central nervous system (Guillemot and Joyner 1993) suggested the existence of other bHLH transcription factors that might compensate for loss of Mash1. In *Drosophila* the proneural gene *atonal* encodes a bHLH factor and participates in formation of chordotonal organs in the peripheral nervous system (Jarman *et al*. 1993). Two murine genes, designated as *Math1* (Akazawa *et al*. 1995) and *Math2* (Shimizu *et al*. 1995), have been found as homologues of *Drosophila atonal*.

Mouse Math2, also named Nex1, *trans*-activates the promoter of its own gene. Expression of the *Nex1* gene coincides with the generation of postmitotic neurons and parallels overt neuronal differentiation and synaptogenesis (Bartholoma and Nave 1994). According to Shimizu *et al*. mouse *Math2* is the target gene of Hes1 and appears to be coexpressed in many regions of the developing nervous system with *Mash1*. It may not only be involved in neural development of embryos but also in the maintenance of subsets of adult mature neural cells. During the 1990s, several genes that encode bHLH proteins that relate to nervous system development and differentiation were cloned from human tissues,

---

such as human *NEUROD*, *NEUROD2* and *NEUROD3* (McCormick *et al.* 1996).

We report here cloning of the human *MATH2* gene, which is highly homologous to the mouse *Math2* gene, from a human foetal brain cDNA library, and its localization in the human chromosome map.

Material and methods

cDNA library and cDNA sequencing: Human foetal brain polyA$^+$ RNA was from Clontech. Double-stranded cDNAs were prepared with the SMART^{TM} PCR cDNA Synthesis Kit (Clontech). cDNAs were digested with *Sfi*I, and fragments of over 500 bp obtained by size fractionation on a Sepharose CL-2B column were used. The pBluescriptII SK(+) vector with a modified MCS was used to directionally clone the cDNAs. Sequencing was performed using either dye-primer or dye-terminator chemistry on an ABI 377 sequencer.

Bioinformatics analysis: Both nucleotide and protein sequence alignments were performed using the BLASTN, BLASTP and BLASTX algorithms on the NCBI web server (http://www.ncbi.nlm.nih.gov/BLAST). GenBank, SwissPro and PDB databases were also used.

Northern blot analysis: The cDNA fragment of 1011 bp (nucleotides 354 to 1365) was amplified by PCR for use as a probe template. *a*-^{32}P-dATP was incorporated into the template using the Prime-a-Gene Labeling System (Promega) according to manufacturer's instructions. A multiple tissue Northern (MTN) blot membrane (Clontech) with Northern blots of eight human tissue mRNAs was hybridized with the cDNA probe according to the manufacturer's user manual. The autoradiograph was scanned using a Cyclone Autoradiography Scanner (Packard). The membrane was then refreshed and probed with a *b*-actin probe for control.

Results and discussion

Cloning and identification of human MATH2 gene

During large-scale sequencing analysis of the human foetal brain cDNA library, we obtained a cDNA clone with an insert of 2176 bp which has an open reading frame (ORF) of 1011 bp (nucleotides 354 to 1365). This ORF encodes a protein of 337 amino acids (figure 1). The 3′-end of the sequence contains the polyA tailing signal and a polyA stretch. There are two in-frame stop codons upstream of the ORF, and the sequence around the proposed initiating ATG is similar to the Kozak conserved sequence, ACCATGG, indicating that the predicted sequence probably represents the full-length protein (Kozak 1987).

```
gtgaaatgctggtgtgcagatttttttcctatcagtctaaccttctgtgttgatgcatgaag    60
gctggtacccacttacagggatgccagatgatcagtgcagaatgaaggtcccaagagaga    120
ggatcacatggctctctctgccctgtgacgtcactagcagatggcatgggtaccagctct    180
ggcagttggcatcaatgtcacttttagagatcaatgagatagtgcagatatacacagat    240
ctagagactccaggagacgatgcgacactcagcctgaaaagatttggaagacccaaatg    300
aaaactgattattgaatgaaattaaaacctaaggtaatttaagattagagagccATGtta    360
                                              M  L
acactaccgtttgatgagtctgttgtaatgccagaatcccagatgtgcagaaagtttct    420
 T  L  P  F  D  E  S  V  V  M  P  E  S  Q  M  C  R  K  F  S
agagaatgcgaggaccagaagcaaattaagaagccagaaagcttttccaaacagattgtc    480
 R  E  C  E  D  Q  K  Q  I  K  K  P  E  S  F  S  K  Q  I  V
cttcgaggaaagagcatcaaaagggcccctggagaagaaaccgagaaagaagaagaggag    540
 L  R  G  K  S  I  K  R  A  P  G  E  E  T  E  K  E  E  E  E
gaagacagggaagaggaagatgaaaatgggttgcctagaaggagggggtcttaggaaaaaa    600
 E  D  R  E  E  E  D  E  N  G  L  P  R  R  R  G  L  R  K  K
aagacaacaaagctgccgattggaaagggtcaagttcaggagacaggaagcgaacgcgcgc    660
 K  T  T  K  L  R  L  E  R  V  K  F  R  R  Q  E  A  N  A  R
gagaggaacaggatgcacggcctcaacgacgctctggacaacttaagaaaagtggtccc    720
 E  R  N  R  M  H  G  L  N  D  A  L  D  N  L  R  K  V  V  P
tgttattctaaaacccagaaactgtccaaaatagagactttacgactggccaaaaactac    780
 C  Y  S  K  T  Q  K  L  S  K  I  E  T  L  R  L  A  K  N  Y
atctgggcacttctgaaattctgagaatcggcaagagaccagatctgctcacattcgtc    840
 I  W  A  L  S  E  I  L  R  I  G  K  R  P  D  L  L  T  F  V
caaaaacttatgcaaaggtctttcccagccaactacaaacttggtggcaggctgcttgcag    900
 Q  N  L  C  K  G  L  S  Q  P  T  T  N  L  V  A  G  C  L  Q
ctcaacgccaggagtttcctgatgggtcaggtggggaggctgcacaccacacaaggtca    960
 L  N  A  R  S  F  L  M  G  Q  G  G  E  A  A  H  H  T  R  S
ccctactctaccttctacccaccctaccacagccctgagctcaccactcccccagggcat    1020
 P  Y  S  T  F  Y  P  P  Y  H  S  P  E  L  T  T  P  P  G  H
gggactcttgataattccaagtccatgaaaccctacaattattgcagtgcgtatgaatcc    1080
 G  T  L  D  N  S  K  S  M  K  P  Y  N  Y  C  S  A  Y  E  S
ttctatgaaagtacttcccctgagtgtgccagccctcagtttgaaggtcccttaagtcct    1140
 F  Y  E  S  T  S  P  E  C  A  S  P  Q  F  E  G  P  L  S  P
ccccaattaactataatgggatattttccctgaagcaagaagaaaccttggactatggt    1200
 P  P  I  N  Y  N  G  I  F  S  L  K  Q  E  E  T  L  D  Y  G
aaaaattacaattacggcatgcattactgtgcagtgccacccaggggtcccccttgggcag    1260
 K  N  Y  N  Y  G  M  H  Y  C  A  V  P  P  R  G  P  L  G  Q
ggtgccatgttcaggttgcccaccgacagccacttcccttacgacttacatctgcgcagc    1320
 G  A  M  F  R  L  P  T  D  S  H  F  P  Y  D  L  H  L  R  S
caatctctcacaatgcaagatgaattaaatgcagttttcataattaatgaggaaaatga    1380
 Q  S  L  T  M  Q  D  E  L  N  A  V  F  H  N
aaataaacagtggtcattcacctcccctgtctaattaagacaaagcagatgcttgtgggc    1440
tgagtaattggcacaactctatctaaggtgtttactagtttctgaagtgtgtttcaacta    1500
ttgtgagaattttctatgtaataataaatctcttttcgtatgagaacttcttttccttc    1560
ctttttgtctgtgaagcactgtgattctgtttctactggaaagatttttttcttttaatt    1620
ttcttttaaacttatttaatttgtttgaacaaggtgtctaagaatatactgttgaataaa    1680
gacatgcacacagcataattcaatgtctattttggttgtacagtaattataaaatgcatg    1740
ttattaaaaccagatgagtaaaatgatgtgtttataattattaggaattatatattatgt    1800
atcttttaaaaattgaaattttaaaatatggagaatatcactaagctgacaatgatattg    1860
ggagatgaatttgcaacaatataaaaagctatgcaattttcttttttattaaggatgaatc    1920
taaatgcattcaattggtaactattcctcttcaagtatttgcagatggggcaacatgata    1980
tttgggtagctgatgtcataattcggggacaattatctcaatcatcaagaaaaaaatttc    2040
agcacttgttttgtattattcagaaattaactgtacaggtttgttatcacaatgtataat    2100
tgtgtttttccaccccacatttttaaggcaattaaatatagatatttccactcgtaaaaa    2160
aaaaaaaaaaaaaaaa
```

Figure 1. Nucleotide sequence of a human foetal brain cDNA cloned insert containing an ORF for human *MATH2* (GenBank accession number AF248954). The predicted amino acid sequence is shown in upper-case in single-letter code. The two stop codons in the 5′-flanking region and the polypeptide portion rich in glutamic acid residues are underlined. The bHLH domain of the protein is shown boxed.

Comparison of the cDNA sequence with the mouse *Math2* sequence (GenBank accession number D44480) revealed 93% homology in nucleotides and 98% homology in amino acids. Therefore we designated the cDNA sequence as human *MATH2* gene. It has been deposited in GenBank with accession number AF248954.

Human MATH2 as a bHLH protein

Human MATH2 contains the bHLH motif and therefore belongs to the bHLH family. Comparison of human MATH2 protein with some other members of the family showed that its bHLH domain has 95% homology with those of human NEUROD (GenBank accession number D82347) and NEUROD2 (GenBank accession number U58681) (figure 2).

Human MATH2 shows high similarity to mouse Math2. The two proteins differ in only four amino acids, and show 100% identity in their bHLH domains. In addition,

both are specifically synthesized in brain. These data suggest that human MATH2 and mouse Math2 may have the same function in the nervous system.

Human MATH2, NEUROD and NEUROD2 contain an amino-terminal domain rich in glutamate residues. This may represent an acidic region indicative of a transcriptional activation domain. Little difference in the bHLH domains of these three proteins implies that they may bind to similar DNA sequences to regulate development and differentiation of the human nervous system. The *NEUROD* gene (*BETA2*) is also expressed in beta cells of pancreas, where it is believed to be an important tissue-specific regulator. It could interact with the E-boxes of insulin gene promoter and thereby regulate transcription of insulin (Naya *et al*. 1995). Human MATH2 and NEUROD have only three amino acid differences in the bHLH domains, and show 100% identity within the loop as well as the second helix. Therefore MATH2 may
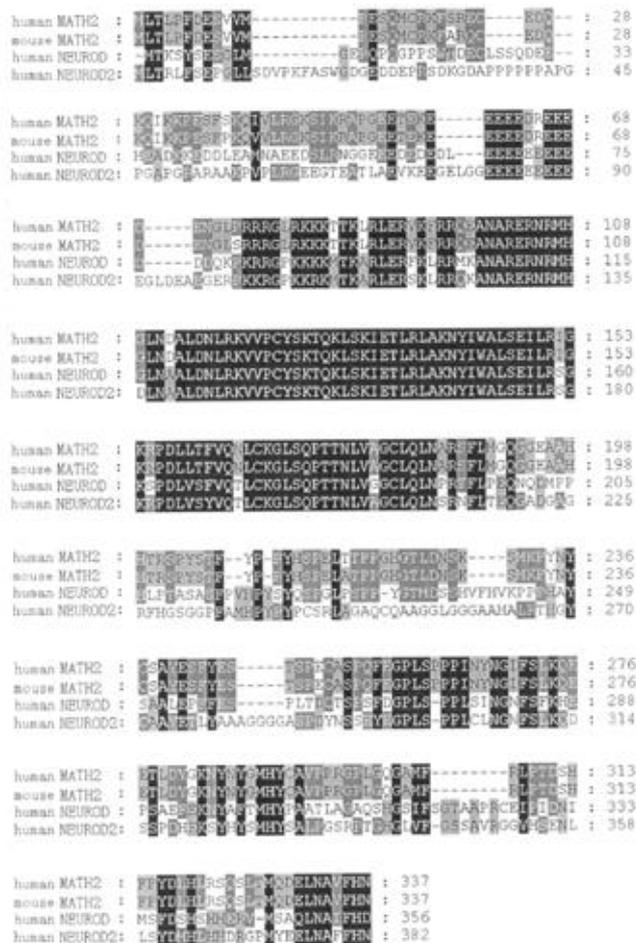

Figure 2. Sequence comparison of human MATH2 protein and homologous proteins mouse Math2 (GenBank D44480), human NEUROD (D82347) and human NEUROD2 (U58681). The portion rich in glutamic acid residues is from position 71 to position 90; the bHLH domain spans positions 123 to 179.
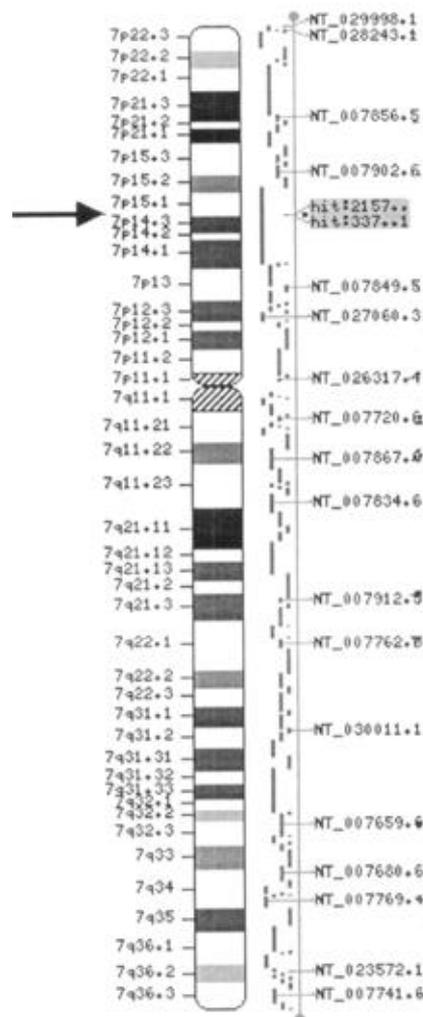
Figure 3. Chromosomal localization of human *MATH2*. The arrow shows the location of human *MATH2* gene on chromosome 7 at p14–15.

Table 1. Nucleotide sequence at exon–intron junctions in human *MATH2*. Exon sequences are shown in upper case and intron sequences in lower case.

3′ Splice acceptor	Exon	Size (bp)	5′ Splice donor	Intron	Size (bp)
(cDNA end)GTGAAATGCTGG	1	333	TTAAAACCTAAGgtaatataaata	1	1297
gcatttttcagGTAATTTAAGAT	2	1821	ATTTCCACTCGT		
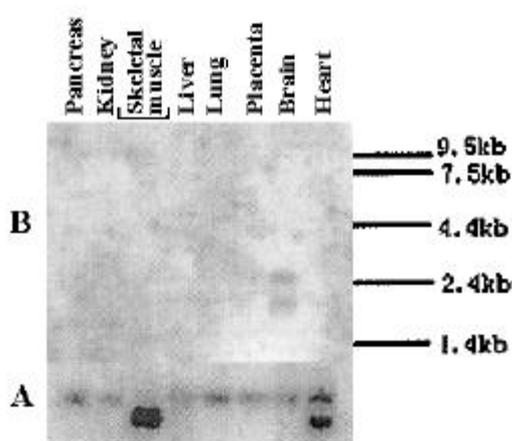
Figure 4. Northern blot analysis of human *MATH2* with an MTN membrane. (A) Positive control, probed with a *b*-actin probe. In both heart and skeletal muscle, there are two forms of *b*-actin mRNA, a 2 kb form and a 1.6–1.8 kb form. (B) Hybridization with the human *MATH2* probe. Two transcripts of human *MATH2* of 1.7 kb and 2.4 kb are seen only in brain.

function by a similar mechanism as NEUROD during regulation.

Chromosomal localization of human MATH2

Comparison of the human *MATH2* cDNA sequence with sequences in the Human Genome Database revealed that our cDNA sequence completely matched two parts (4662967–4662630 and 4661333–4659507) of the BAC (bacterial artificial chromosome clone) sequence AC006380 (contig NT_007819), which is located on the human chromosome map at 7p14–15 (figure 3). The sequence comparison also showed that there is an intron of 1.3 kb with 5′ and 3′ splice sites between nucleotides 333 and 334 of the cDNA sequence (table 1).

Northern blots

Northern blot analysis revealed two *MATH2* transcripts of 1.7 kb and 2.4 kb, which were specifically present in brain among the eight tissues examined (figure 4). Similar results were obtained in Northern blots for mouse *Math2* (Shimizu *et al*. 1995). The size of the 2.4 kb transcript is consistent with the length of human *MATH2*.

The band at 2.4 kb is the proper transcript of human *MATH2*, while the origin of the band at 1.7 kb is unknown. We suppose that there are different splicing modes of human *MATH2* or that there are other homologous genes in the human nervous system.

References

Akazawa C., Ishibashi M., Shimizu C., Nakanishi S. and Kageyama R. 1995 A mammalian basic helix–loop–helix factor structurally related to the product of *Drosophila* proneural gene *atonal* is a positive transcriptional regulator expressed in the developing nervous system. *J. Biol. Chem.* **270**, 8730–8738.

Bartholoma A. and Nave K. A. 1994 Nex-1: a novel brain-specific basic helix–loop–helix protein with autoregulation and sustained expression in mature cortical neurons. *Mech. Dev.* **48**, 217–228.

Casarosa S., Fode C. and Guillemot F. 1999 MASH1 regulates neurogenesis in the ventral telencephalon. *Development* **126**, 525–534.

Guillemot F. and Joyner A. L. 1993 Dynamic expression of the murine achaete-scute homologue MASH1 in the developing nervous system. *Mech. Dev.* **42**, 171–185.

Ishibashi M., Moriyoshi K., Sasai Y., Shiota K., Nakanishi S. and Kageyama R. 1994 Persistent expression of basic helix–loop–helix factor HES1 prevents mammalian neural differentiation in the central nervous system. *EMBO J.* **13**, 1799–1805.

Jarman A. P., Grau Y., Jan L. Y. and Jan Y. N. 1993 *Atonal* is a proneural gene that directs chordotonal organ formation in the *Drosophila* peripheral nervous system. *Cell* **73**, 1307–1321.

Kozak M. 1987 An analysis of 5′-noncoding sequence from 699 vertebrate messenger RNAs. *Nucl. Acids Res.* **15**, 8125–8148.

Lo L. C., Johnson J. E., Wuenschell C. W., Saito T. and Anderson D. J. 1991 Mammalian achaete-scute homolog 1 is transiently expressed by spatially restricted subsets of early neuroepithelial and neuro crest cells. *Genes Dev.* **5**, 1524–1537.

McCormick M. B., Tamimi R. M., Snider L., Asakura A., Bergstrom D. and Tapscott S. J. 1996 *NEUROD2* and *NEUROD3*: Distinct expression patterns and transcriptional

activation potentials within the *NEUROD* gene family. *Mol. Cell. Biol.* **16**, 5792–5800.

Naya F. J., Stellrecht C. M. and Tsai M. J. 1995 Tissue-specific regulation of the insulin gene by a novel basic helix–loop–helix transcription factor. *Genes Dev.* **9**, 1009–1019.

Sasai Y., Kageyama R., Tagawa Y., Shigemoto R. and Nakanishi S. 1992 Two mammalian basic helix–loop–helix factors structurally related to *Drosophila* hairy and Enhancer of split. *Genes Dev.* **6**, 2620–2634.

Shimizu C., Akazawa C., Nakanishi S. and Kageyama R. 1995 MATH2, a mammalian basic helix–loop–helix factor structurally related to the product of *Drosophila* proneural gene *atonal*, is specifically expressed in the nervous system. *Eur. J. Biochem.* **229**, 239–248.