Development 127, 1185-1196 (2000)
Printed in Great Britain © The Company of Biologists Limited 2000
DEV9700

Autoregulation and multiple enhancers control *Math1* expression in the developing nervous system

Amy W. Helms[1], Andrew L. Abney[1], Nissim Ben-Arie[2], Huda Y. Zoghbi[3] and Jane E. Johnson[1,*]

[1]Center for Basic Neuroscience, University of Texas Southwestern Medical Center, 5323 Harry Hines Blvd, Dallas, TX 75235, USA
[2]Department of Cell and Animal Biology, Alexander Silberman Institute of Life Sciences, The Hebrew University of Jerusalem, Givat-Ram, Jerusalem, Israel, 91904
[3]Department of Molecular and Human Genetics, Howard Hughes Medical Institute, Baylor College of Medicine, Houston, TX 77030, USA
*Author for correspondence (e-mail: Jane.Johnson@email.swmed.edu)

Accepted 17 December 1999; published on WWW 21 February 2000

SUMMARY

Development of the vertebrate nervous system requires the actions of transcription factors that establish regional domains of gene expression, which results in the generation of diverse neuronal cell types. MATH1, a transcription factor of the bHLH class, is expressed during development of the nervous system in multiple neuronal domains, including the dorsal neural tube, the EGL of the cerebellum and the hair cells of the vestibular and auditory systems. MATH1 is essential for proper development of the granular layer of the cerebellum and the hair cells of the cochlear and vestibular systems, as shown in mice carrying a targeted disruption of *Math1*. Previously, we showed that 21 kb of sequence flanking the *Math1*-coding region is sufficient for *Math1* expression in transgenic mice. Here we identify two discrete sequences within the 21 kb region that are conserved between mouse and human, and are sufficient for driving a *lacZ* reporter gene in these domains of *Math1* expression in transgenic mice. The two identified enhancers, while dissimilar in sequence, appear to have redundant activities in the different *Math1* expression domains except the spinal neural tube. The regulatory mechanisms for each of the diverse *Math1* expression domains are tightly linked, as separable regulatory elements for any given domain of *Math1* expression were not found, suggesting that a common regulatory mechanism controls these apparently unrelated domains of expression. In addition, we demonstrate a role for autoregulation in controlling the activity of the *Math1* enhancer, through an essential E-box consensus binding site.

Key words: bHLH, Transcription factor, Cerebellum, EGL, Ear, Whisker vibrissae, Sensory epithelium, Neural development, Merkel cell

INTRODUCTION

The molecular mechanisms underlying development of the vertebrate nervous system are complex and largely undefined. In the last few years, members of the basic-helix-loop-helix (bHLH) family of transcription factors have come to the forefront as playing an essential role in neural development (Kageyama et al., 1995; Lee, 1997). From *Drosophila* to mammals, bHLH molecules have been implicated in both positive and negative regulatory roles ranging from neural determination, or proneural gene function (Cabrera and Alonso, 1991; Henrique et al., 1997; Jan and Jan, 1993; Ma et al., 1996), to the process of differentiation from proliferating neural precursor to postmitotic neuron (Guillemot et al., 1993; Helms and Johnson, 1998; Johnson et al., 1990; Lee et al., 1995; Shimizu et al., 1995).

The *Drosophila* proneural genes of the *achaete-scute* complex (ac-sc) and *atonal* are required for development of specific subsets of neurons (Campos-Ortega and Jan, 1991; Jan and Jan, 1993; Jarman et al., 1993, 1994, 1995). Their vertebrate counterparts include the mammalian *achaete-scute* homolog *Mash1* and multiple *atonal* homologs, which include *Math1*, *ngn1*, *ngn2*, *NeuroD*, *Math2/Nex-1* and *NeuroD2* (Akazawa et al., 1995; Ben-Arie et al., 1996; Johnson et al., 1990; Lee et al., 1995; Ma et al., 1996; McCormick et al., 1996; Shimizu et al., 1995). These bHLH transcription factors are expressed transiently in discrete patterns during early development of the vertebrate nervous system, when neural specification and early differentiation are taking place. MASH1, MATH1, NGN1 and NGN2 are essential for neural development. Development of both the central and peripheral nervous system is severely disrupted in *Mash1* mutants (Casarosa et al., 1999; Horton et al., 1999; Torii et al., 1999; Tuttle et al., 1999), development of the cerebellum and sensory epithelium of the ear is impaired in the *Math1* mutant (Ben-Arie et al., 1997; Bermingham et al., 1999), and development of cranial ganglia is disrupted in *ngn1* and *ngn2* mutants (Fode et al., 1998; Ma et al., 1998). Because of the essential role of bHLH transcription factors during development, and the discrete nature of their temporal and spatial expression

patterns, it is likely that precise control of gene expression is necessary for proper neural development. Identification of upstream regulators of bHLH expression will be key in deciphering patterning within the nervous system and the mechanism of commitment of precursors to specific neural fates.

Much has been achieved toward an understanding of the regulation of *atonal* and the *achaete-scute* complex in *Drosophila*. Enhancer elements that regulate *ac-sc* expression span a 90 kb regulatory region, and some of these enhancers are shared between discrete genes of the complex (Gomez-Skarmeta et al., 1995). A number of molecules have been shown to regulate specific aspects of the *ac-sc* expression pattern, including the homeodomain-containing genes, *araucan* and *caupolican*, negative regulators such as *hairy*, *emc*, *pannier* and *u-shaped*, and positive regulation by *achaete* itself (Cubadda et al., 1997; Gomez-Skarmeta et al., 1996; Leyns et al., 1996; Martinez and Modolell, 1991; Ramain et al., 1993; Skeath and Carroll, 1991; van Doren et al., 1992, 1994). In vertebrates, multiple sequence elements directing *Mash1* expression in the CNS have been identified in a 1.2 kb region that is approximately 7 kb upstream of the *Mash1*-coding region (Verma-Kurvari et al., 1996, 1998) and, it appears that regulation of the entire complement of *Mash1* expression is highly complex, much like its *Drosophila* counterpart, the AS-C. By comparison, the regulation of *atonal* is less complex. The regulatory regions governing *atonal* expression are arranged in modular elements 9.3 kb upstream and 5.8 kb downstream of the *atonal*-coding region. Some, but not all of these modular elements are Atonal-dependent, meaning that Atonal autoregulates in multiple domains of expression (Sun et al., 1998).

In this study, we have focused on regulation of the vertebrate *atonal* homolog, *Math1*. It has been shown that *Math1* is expressed in the embryonic vertebrate nervous system in the developing spinal cord and hindbrain, and the developing sensory epithelium in the ear. Hindbrain expression persists in the external granular layer (EGL) of the developing cerebellum until well after birth (Akazawa et al., 1995; Ben-Arie et al., 1996). *Math1*-expressing neural precursors in the developing spinal cord become dorsal commissural interneurons (Helms and Johnson, 1998) and *Math1*-expressing cells in the EGL of the cerebellum become granule cells (Ben-Arie et al., 1997). In the developing ear, MATH1 is expressed in progenitors to the hair cells in the cochlea and semicircular canals. A general feature of MATH1 expression is that it is transient and is expressed in proliferating progenitor populations but is extinguished soon after cells become postmitotic and begin to express markers characteristic of specific neuronal types (Helms and Johnson, 1998). Mice mutant for *Math1* die at birth and exhibit a complete absence of cerebellar granule cells (Ben-Arie et al., 1997). In addition, *Math1* mutant mice are devoid of hair cells that line the cochlea and semicircular canals, which are essential for both hearing and proper balance (Bermingham et al., 1999).

Because of its complex spatial and temporal pattern of expression, and essential function during cerebellar and hair cell development, it is likely that precise regulation of *Math1* expression is crucial for proper neuronal development. Here we delineate two enhancers located 3′ of the *Math1*-coding region, which are sufficient for driving expression of *Math1* in its diverse expression domains, including a previously unknown expression domain for *Math1* in mechanosensory cells of the developing whisker vibrissae. We have identified *cis*-acting elements that are required for *Math1* expression as a first step to discover the signaling pathways that control early neuronal development through bHLH transcription factors. In addition, within these elements, we have identified an essential E-box and a role for MATH1 in autoregulation.

MATERIALS AND METHODS

Transgene construction

Two overlapping *Math1*-containing genomic clones were isolated from a Stratagene λ129 mouse library using a probe from the bHLH region of *Math1*. The assembly of transgene 1 (the *Math1/lacZ* transgenic line) has been described previously (Helms and Johnson, 1998). It contains 15 kb 5′ and 6 kb 3′ of the *Math1*-coding region. Transgene 2 contains a 15 kb 5′ *Sph*I fragment fused to the *lacZ* reporter and *SV40* polyadenylation site derived from pnLacF (Mercer et al., 1991) such that the *lacZ* gene is in frame with the *Math1* ATG. Transgenes 3-6 also place the *lacZ*-coding region in frame with the *Math1* ATG and retains both the 5′ and 3′ UTRs as described for transgene 1. Relevant restriction enzyme sites defining the ends of the *Math1* sequence included in each transgene are indicated in Fig. 3A.

Transgenes 7-17 place sequences from the 3′ end of the *Math1* gene 5′ of the heterologous promoter construct BGZA (Yee and Rigby, 1993). BGZA contains the β-globin basal promoter, the *lacZ*-coding region and *SV40* polyadenylation sequences. For assembly of transgenes 7 and 8, the restriction fragments used are indicated in Fig. 3A. Transgenes 9-17 were constructed using available restriction enzyme sites and PCR (specific details available upon request; Fig. 5A). All constructs generated by PCR were sequenced across the PCR region.

To generate an epitope-tagged *Math1* transgene (transgene 5), an epitope from the *Drosophila* Bride of Sevenless (BOSS) protein was used (Krämer et al., 1991). The BOSS epitope used contained the following amino acid sequence from the N-terminal region of BOSS: TSPTKKSAPLRITKPQPTS. The epitope-tagged construct was generated by creating an oligonucleotide pair that encoded the 19 amino acid BOSS epitope, flanked by *Sph*I sites. The following oligonucleotides were used. BO1 5′ CAA CCA GCC CTA CCA AGA AGA GCG CCC CTC TGA GAA TCA CCA AGC CTC AGC CTA CCA GCG CAC ATG 3′ and BO2 5′ TGC GCT GGT AGG CTG AGG CTT GGT GAT TCT CAG AGG GGC GCT CTT CTT GGT AGG GCT GGT TGC ATG 3′. These oligonucleotides were annealed and inserted into the *Math1*-coding region at an *Sph*I site located 7 amino acids downstream of the *Math1* start ATG. The ^{BOSS}*Math1* reporter construct injected contains the same *Math1*-flanking DNA as transgene 4, with ^{BOSS}*Math1* replacing *lacZ* (Fig. 3A).

Transgene fragments were isolated from vector sequences and prepared for injection as previously described (Helms and Johnson, 1998).

Generation and analysis of transgenic mice

Transgenic mice were generated by standard procedures (Hogan et al., 1986) using fertilized eggs from B6D2F1 (C57Bl/6×DBA) crosses. Transgenic embryos were either assayed for transgene expression in founder embryos (G_0) or in embryos obtained from an established transgenic line. Transgenic founder animals were identified by Southern or dot-blot analysis using tail DNA (established lines) or yolk sac DNA (G_0 embryos). The transgene integration analysis for transgene 1 has been previously described (Helms and Johnson, 1998). Transgene integration analysis for transgene 4 was characterized by Southern blot and indicated that both fragments had recombined and integrated correctly (data not shown). A 2.5 kb *Pvu*II fragment from *lacZ* was used to probe for transgenes containing *lacZ*

and an oligonucleotide containing the BOSS epitope sequence and/or a 600 bp *Sph*I-*Bgl*II fragment containing the *Math1*-coding region were used to probe for transgenes containing the ^{BOSS}*Math1* reporter. Founder animals were outbred with B6D2F1 animals for all studies. Embryos were staged based on the assumption that copulation occurred at E0, halfway through the dark cycle.

Genotypic analysis of the *Math1* mutant mice crossed into the J2G transgenic strain (Tg 6) was carried out by Southern blot. Yolk-sac DNA was prepared and digested with *Eco*RI and *Bam*HI separately. Southern blots with *Eco*RI-digested DNA were hybridized with a 400 bp *Spe*I probe derived from the region 1 kb 5′ of the *Math1* start site and outside the *Math1/lacZ* transgene yielding a 10 kb band from the wild-type allele and a 4 kb band from the mutant allele. Southern blots with *Bam*HI-digested DNA were hybridized with a 2.5 kb *Pvu*II probe derived from the coding region of *lacZ* yielding a 3 kb band if the transgene is present.

β-galactosidase staining of embryos

Staged transgenic embryos and postnatal brains were dissected in cold PBS and fixed for 30-60 minutes, depending on the age of the embryo, in 4% formaldehyde pH 7.2 at room temperature. Whole-mount β-gal staining of the embryos was carried out as described (Verma-Kurvari et al., 1996). For analysis of thin sections, fixed embryos were sunk in 30% sucrose overnight at 4°C, embedded in OCT and cryosectioned at 30 μm. Slides were stained for β-gal activity at 35°C overnight in 1 mg/ml 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside (X-gal), 5 mM $K_3Fe(CN)_6$, 5 mM $K_4Fe(CN)_6$, 5 mM $MgCl_2$ in PBS.

Immunocytochemistry and Merkel cell detection

Immunocytochemistry was carried out as previously described (Helms and Johnson, 1998). Rabbit polyclonal antibodies against MATH1 were used to detect MATH1 protein (Helms and Johnson, 1998). Rabbit polyclonal antibody NN-1 and mouse monoclonal antibody BOSS-1 (Krämer et al., 1991) were used to detect the BOSS epitope. Merkel cells were detected using RG 53 (ICN pharmaceuticals), a monoclonal antibody to cytokeratin-18, which is a marker of Merkel cells (Airaksinen et al., 1996). Immunofluorescence imaging was carried out on a Bio-Rad MRC 1024 confocal microscope.

Electrophoretic mobility shift assay

The electrophoretic mobility shift assay (EMSA) was carried out using in vitro transcribed and translated MATH1 and E12 proteins. The in vitro transcription/translation products were generated using the TnT T7 or Sp6 Coupled Reticulocyte Lysate System (Promega) using plasmids containing the *Math1*-coding region (pMATH1c) and the E12-coding region (pCITE.E12) (Black et al., 1998) as templates. Protein synthesis was confirmed on an SDS-PAGE gel. For reactions containing both MATH1 and E12, proteins were pre-bound at 37°C for 20 minutes before binding reactions were set up. The following oligonucleotides were used for the EMSA: Wild type: #1 5′ CAT GGC ACG CGC TGT CAG CTG GTG AGC GCA C 3′ and #2 5′ CAT GGT GCG CTC ACC AGC TGA CAG CGC GTG C 3′.

Mutant: #1 5′ CAT GGC ACG CGC TGT ATT CTG GTG AGC GCA C 3′ and #2 5′ CAT GGT GCG CTC ACC AGA ATA CAG CGC GTG C 3′.

Other E box: #1 5′ CAT GGG CTC ATT CCC CAT ATG CCA GAC CAC GC 3′ and #2 5′ CAT GGC GTG GTC TGG CAT ATG GGG AAT GAG CC 3′.

Oligonucleotides were annealed, end-labeled with γ^{32}P-dATP, and purified on a Bio-Spin 6 column (Bio-Rad). The binding reactions were carried out as follows: 1× gel shift buffer (100 mM Hepes pH 7.6, 250 mM KCl, 5 mM DTT, 5 mM EDTA, and 25% glycerol), 1 μg poly dI-dC, 28,000 cts/minute labeled probe (~0.25 ng) and 4 μl of the appropriate reticulocyte lysate. 100× molar ratio of cold annealed oligonucleotide was used in the competition experiments. Binding reactions were carried out at room temperature for 15 minutes and were run on a non-denaturing polyacrylamide gel.

Isolation of *Hath1* genomic clones

The *Hath1*-containing genomic clones were isolated from a human genomic library (Clontech) by low-stringency hybridization using a 333 bp *Sac*II-*Bgl*II probe derived from the coding region of *Math1*. Lambda clones containing sequences that hybridized with the *Math1*-coding probe were isolated using the LambdaSorb purification kit (Stratagene) and mapped using the *Sac*II-*Bgl*II *Math1*-coding probe and the 1.7 kb *Sac*II-*Sma*I enhancer probe (from transgene 9). Two mostly overlapping lambda clones containing both the *Hath1*-coding and enhancer regions were sequenced and compared with the corresponding regions of *Math1*. Sequence conservation identified the enhancers in the *Hath1* sequence located 3.4 kb 3′ of the *Hath1*-coding region (see diagram Fig. 5A). The *Math1* and *Hath1* enhancer sequences have been submitted to GenBank (accession numbers: AF218258 and AF218259, respectively).

RESULTS

Math1/lacZ transgenic lines recapitulate *Math1* expression in multiple domains

Previously, we generated a line of transgenic mice that expresses *lacZ* under the control of *Math1*-flanking DNA sequences (Helms and Johnson, 1998). This transgenic strain was generated by co-injection of three overlapping fragments of DNA that together contain 15 kb 5′ and 6 kb 3′ *Math1* sequence flanking the coding region (transgene 1). At embryonic stages E10 through E11.5, restriction of transgene expression to the neural tube was shown to be similar to the endogenous pattern of *Math1* expression in both its timing and rostrocaudal boundaries of expression (Helms and Johnson, 1998).

To determine whether the 21 kb of regulatory sequence governs expression in all *Math1* domains, we compared *lacZ* expression with endogenous *Math1* expression at several different stages that reflect these domains. In addition to the dorsal neural tube expression pattern of *Math1*, three other domains of expression have been previously reported. These domains include the cranial ganglia at E9.5, the EGL of the cerebellum from rhombic lip stages (E13-E15) through postnatal development, and the developing hair cells of the cochlea and semicircular canals (Akazawa et al., 1995; Bermingham et al., 1999). *Math1* and *lacZ* expression in the EGL of the cerebellum was analyzed on the day of birth, P0. Frozen sections were either immunolabeled with the MATH1 antibody or stained with X-gal to reveal β-gal activity. Both MATH1 and β-gal were localized to the EGL in an identical spatial and temporal pattern, demonstrating the presence of appropriate regulatory sequence information present in the transgene for EGL expression (Fig. 1A,B). At E14.5, MATH1 protein is expressed in the sensory epithelium of the developing ear in the developing cochlea (Bermingham et al., 1999) and semicircular canals (Fig. 1C). In an adjacent section, localization of β-gal activity mimics this domain of expression (Fig. 1D). In contrast, transgene expression in the cranial ganglia at E9.5 has never been detected (data not shown). While previous reports using mRNA in situ hybridization demonstrate expression in the cranial ganglia (Akazawa et al., 1995), we have been unable to confirm this expression domain using the MATH1 antibody (data not shown). Thus, protein

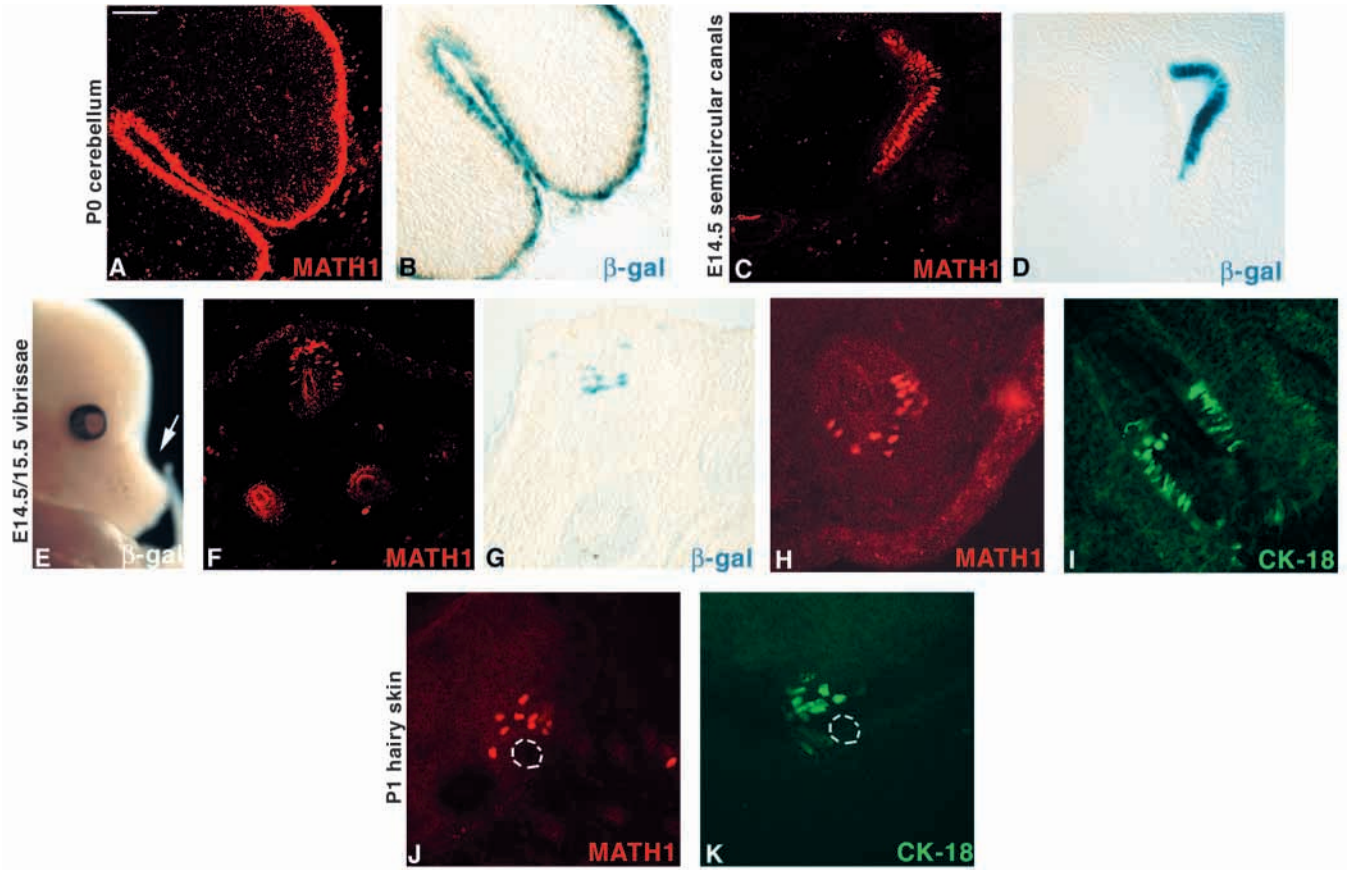

Fig. 1. Expression of the *Math1/LacZ* transgene recapitulates endogenous *Math1* expression in multiple domains. MATH1 immunofluorescence (A,C,E,F,H,J), staining for β-gal activity (B,D,E,G), or cytokeratin 18 (CK-18) (I,K) in the transgenic line J2C (Tg 1). (A,B) Sagittal sections of the postnatal day 0 cerebellum. (C,D) transverse sections of E14.5 developing semicircular canal ampulla. (E) Whole-mount β-gal staining of E14.5 embryo showing whisker vibrissae (arrow), (F-I) Transverse sections of E14.5 or E15.5 whisker vibrissae. (J, K) Epidermal Merkel cells in postnatal day 1 hairy skin. The touch domes are outlined with a dashed line. The β-gal staining in the transgenic embryos mimics the endogenous expression pattern of MATH1 shown by immunofluorescence except in the Merkel cells of hairy skin. Scale bar 50 μm in A-D, F and G, 25 μm in H-K and 1.4 mm in E.

expression may be non-existent or below the level of detection with the antibody.

An additional domain of *Math1* expression not previously reported was revealed during examination of E14.5 *Math1/lacZ* transgenic embryos. β-gal activity was detected in patches of cells on the snout in whole mount (Fig. 1E). Localization of β-gal activity in cryosections through this region suggests the expression represents developing mechanoreceptor cells in the whisker vibrissae called Merkel cells (Fig. 1G). In an adjacent section, immunolabeling with the MATH1 antibody verified endogenous *Math1* expression in this domain (Fig. 1F). To confirm the identity of the MATH1-expressing cells as the mechanosensory cells in the whisker vibrissae, the localization of MATH1 staining was compared in adjacent sections with a Merkel-cell-specific marker, cytokeratin-18, at E15.5 (Airaksinen et al., 1996).

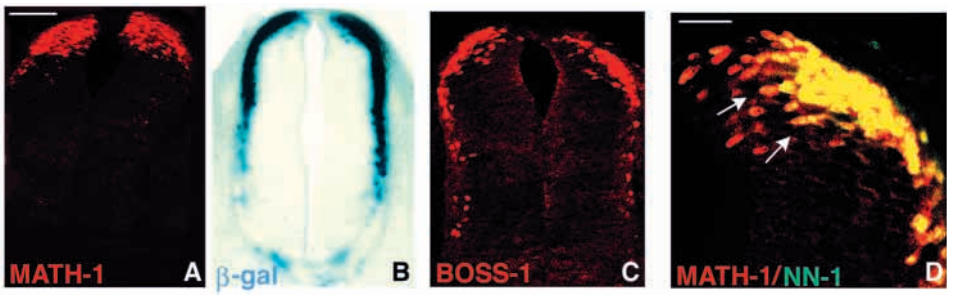

Fig. 2. Expression of the epitope-tagged *Math1* reporter gene in the neural tube. Cross sections through the neural tube at the upper limb level of E10.5 embryos. (A) Anti-MATH1 antibody showing the endogenous *Math1* expression domain. (B) *Math1/lacZ* transgenic line 1 stained for β-gal activity to demonstrate the persistence of transgene expression in differentiated dorsal commissural interneurons. (C) Anti-BOSS-1 antibody reveals persistence of the epitope-tagged *Math1* reporter in differentiated cells as well. (D) Double-label immunofluorescence with anti-MATH1 (red) and anti-NN-1 (green), an anti-BOSS-1 monoclonal antibody. Cells expressing the *bossMath* transgene are seen as yellow (arrows). Cells expressing only the endogenous gene are red. Scale bar, 60 μm in A-C and 25 μm in D.

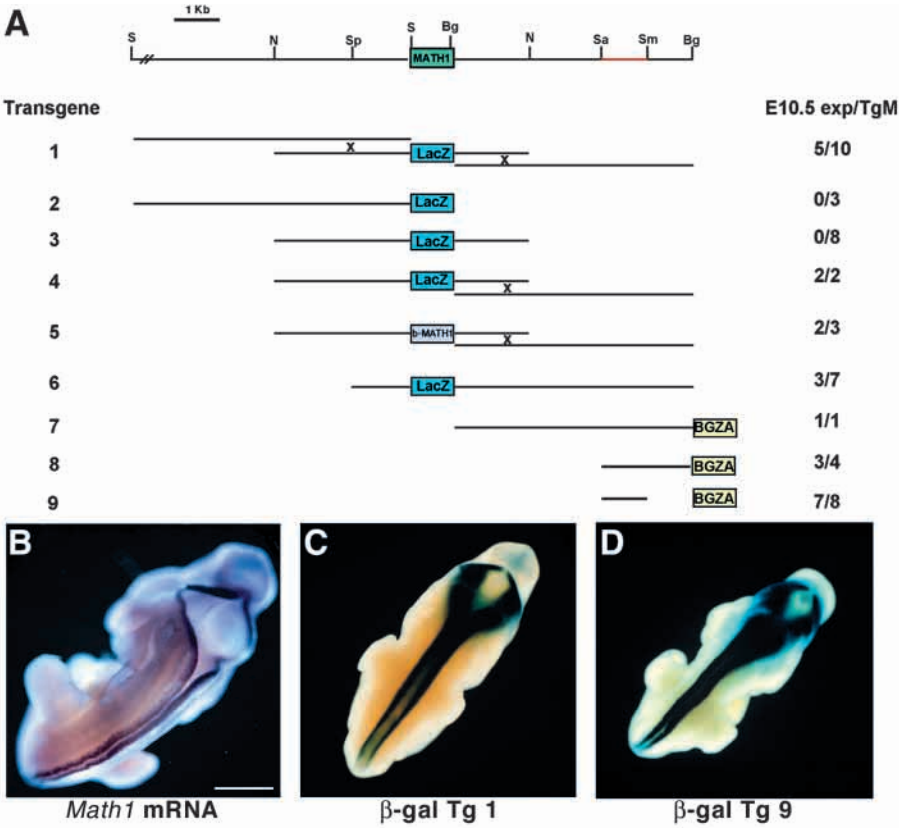

Fig. 3. Identification of a 1.7 kb enhancer element that drives expression of *lacZ* in the spinal neural tube at E10.5. (A) Diagrammatic representation of the genomic structure of the *Math1* gene, and the *Math1/LacZ* fusion constructs used for injection. The green box represents the *Math1*-coding region and the red line represents the location of the identified enhancer relative to the coding region. S, *Sph*I; N, *Nco*I; Sp, *Spe*I; Bg, *Bgl*II; Sa, *Sac*II; Sm, *Sma*I. Constructs 1-6 contain the *lacZ*-coding sequence flanked by *Math1* sequences on the 5′ and 3′ ends. Constructs 7-9 contain *Math1* regulatory sequences fused 5′ of a heterologous *β-globin promoter/lacZ* reporter (BGZA). Most embryos were analyzed as founder embryos (G_0). Data is presented as the number of embryos expressing the transgene in the *Math1*-specific pattern at E10.5 per the number of transgenic embryos identified by DNA analysis (TgM). (B) Whole-mount in situ hybridization of an E10.5 embryo with a *Math1* antisense probe demonstrates the endogenous expression in the dorsal neural tube from the midbrain/hindbrain boundary caudally to the tail. (C, D) Dorsal view of a representative whole-mount β-gal-stained transgenic embryo from transgenic lines 1 and 9 showing that the expression of the reporter gene reflects *Math1*-specific expression. Scale bar 1.2 mm.

MATH1 and CK-18 appear to be expressed in similar cells in the vibrissae (Fig. 1H,I). Because Merkel cells are also found adjacent to touch domes in hairy skin, we investigated whether MATH1 and the *Math1/lacZ* transgene are expressed in this additional domain. MATH1 and the Merkel-specific marker, CK-18, appear to co-localize in the hairy skin in the Merkel cells surrounding the touch dome as demonstrated in adjacent sections of postnatal day 1 skin (Fig. 1J,K). These data confirm the expression in developing Merkel cells in both the whisker vibrissae and in touch domes in hairy skin. In contrast, *lacZ* expression was not detected in the Merkel cells of hairy skin (data not shown). The absence of the *Math1/lacZ* transgene expression in hairy skin is the only MATH1 domain not represented by transgene expression. Taken together, these data demonstrate that the expression of the *Math1/lacZ* transgene containing 15 kb 5′ and 6 kb 3′ *Math1*-flanking sequence, correlates precisely with most domains of *Math1* expression that we have detected using the MATH1 antibody.

Persistence of transgene expression in the dorsal neural tube reflects the lack of a control element specific to this expression domain

Previously we reported that the *Math1/lacZ* transgene expression in the dorsal neural tube aberrantly persisted into differentiated neurons derived from the MATH1-expressing progenitor cells (Helms and Johnson, 1998). The reason for this discrepancy in expression pattern between the reporter gene and the endogenous gene could be due to the greater stability of the reporter *lacZ* mRNA or β-gal protein. Alternatively, this discrepancy in expression could reflect missing regulatory information in the sequence contained in the transgene. To distinguish between these two possibilities, we generated a transgenic line using the *Math1*-flanking sequences that direct expression in the dorsal neural tube, but replaced the *lacZ* reporter from transgene 4 (Tg 4) with a reporter encoding an epitope-tagged MATH1 protein (Fig. 3A, Tg 5). This reporter should be similar in stability to endogenous *Math1* mRNA and protein.

We used the N-terminal 19 amino acids of the *Drosophila* Bride of Sevenless (BOSS) protein as an epitope tag since antibodies to this epitope have been characterized (Krämer et al., 1991). The 19 amino acid BOSS epitope was inserted in-frame into the N terminus of the MATH1 protein 7 amino acids downstream of the start ATG. Transverse sections through the neural tube of E10.5 embryos containing the epitope-tagged *Math1* were analyzed for transgene expression. The two transgenic founder embryos (G_0) that expressed the transgene did so in a pattern mimicking the *Math1/lacZ* line (compare Fig. 2A-C). The transgene-expressing cells overlap the endogenous MATH1 protein in the most dorsal region of the neural tube (Fig. 2C,D). However, like the *Math1/lacZ* transgenes, the ^{BOSS}MATH1-expressing cells are also found in more ventral/lateral positions than the endogenous MATH1 (compare Fig. 2A-C). These data support the conclusion that the persistence of transgene expression is due, at least in part, to missing regulatory information in the transgene rather than mRNA or protein stability alone. One caution in this interpretation, however, is that we cannot rule out an effect on protein stability due to the 19 amino acid epitope. It is important to note that this flaw in transgene regulation appears specific to the dorsal neural tube expression domain since this

persistence in reporter gene expression is not apparent in the other domains of transgene expression (Fig. 1).

Math1 cis-acting regulatory elements are located within a 1.7 kb enhancer region 3′ of the *Math1*-coding region

To define more precisely the sequences responsible for regulation of *Math1* expression, and to determine if separate cassettes are required for expression in different domains, we tested multiple deletion constructs in a transient transgenic mouse assay. Transgenic G_0 embryos were identified by Southern analysis of yolk sac DNA and assayed for *Math1*-specific β-gal expression in the dorsal neural tube at E10.5. Transgene expression was seen only in transgenic embryos containing the 6 kb sequence 3′ of the *Math1*-coding region (Fig. 3A, Tg 1, 4 and 6), but not in transgenic lines containing 15 kb sequence 5′ and 1.5 kb sequence 3′ of the coding region (Fig. 3A, Tg 2, 3). The 6 kb 3′ sequence, and two smaller sequences within the 6 kb (2.6 kb and 1.7 kb) were tested for enhancer activity on a heterologous promoter BGZA (see Methods). As predicted, the 6 kb 3′ sequence was sufficient to direct *lacZ* expression from the heterologous promoter in a *Math1*-specific pattern (Fig. 3A, Tg 7). The smaller 2.6 and 1.7 kb sequences functioned just as well as the 6 kb sequence to direct expression to the dorsal neural tube at E10.5 (Fig. 3A, Tg 8, 9). All expressing transgenes (Tg 1, 4, 6-9) gave a pattern of *lacZ* expression reflecting the *Math1* expression domain in the neural tube at E10.5 (Fig. 3B-D). The enhancer activity of the 1.7 kb fragment found 3.4 kb 3′ of the coding region recapitulates transgene expression detected from 21 kb of *Math1*-flanking sequence.

Regulatory sequences for the diverse domains of *Math1* expression co-localize in the 1.7 kb enhancer sequence 3′ of the gene

The 1.7 kb enhancer is sufficient to direct expression of a heterologous reporter gene to the dorsal neural tube at E10.5. Regulation of genes that are expressed in multiple tissues often have distinct enhancers specific to each different domain of expression (Black and Olson, 1998; Manzanares et al., 1999; Naya et al., 1999; Sun et al., 1998; Zhang et al., 1997). To determine if the regulatory elements for the different domains of *Math1* expression were localized to different enhancer sequences, we assessed the ability of transgene 9, containing the 1.7 kb enhancer, to direct *lacZ* expression to *Math1* expression domains at E14.5 and P0. A stable line of transgenic mice was generated so that transgene expression at each developmental stage could be assayed. *Math1*-specific *lacZ* expression was observed at E10.5 in the dorsal neural tube as had been seen in G_0 embryos in the transient transgenic assay (compare Figs 2B and 4A). In addition, at E14.5 and P0, transgene expression was restricted to the EGL of the cerebellum (Fig. 4B), and the developing ear and whisker vibrissae (Fig. 4C,D) in the precise patterns previously seen. Thus, the 1.7 kb enhancer contains elements sufficient for specific expression in several *Math1* expression domains.

Sequence comparison of *Math1* and *Hath1* identifies two conserved regions within the functional 1.7 kb enhancer

To determine whether sequences controlling the regulation of *Math1* expression are conserved between species, we cloned and sequenced the genomic region surrounding the human *Math1* ortholog, *Hath1* (Ben-Arie et al., 1996). A human genomic clone was isolated that hybridized to a 333 bp *Sac*II-*Bgl*II probe from the *Math1*-coding region. A 6.5 kb *Sal*I fragment from the genomic clone hybridized with both the coding region probe and a probe containing the 1.7 kb *Math1* enhancer. This 6.5 kb fragment was subcloned and sequenced. Comparison of the *Hath1* genomic sequence to the 6 kb sequence 3′ of the *Math1* gene identified two discrete regions of homology outside the transcribed region. A 561 bp sequence, termed enhancer A, is 92% identical between mouse and human, and enhancer B, 544 bp, is 87% identical. The two enhancers are separated by non-conserved sequence of different lengths, 544 bp in human and 413 bp in mouse (Fig. 5A). Sequences outside the homologies diverged completely except in the transcribed regions of the genes. This level of conservation between the regulatory sequences (87-92%) is even greater than the conservation in the regions of the genes encoding the proteins (85%). All but 133 bp of the sequence identified by conservation is contained in the 1.7 kb enhancer demonstrated to drive *Math1*-specific expression in transgenic mice (Fig. 5A). Deletion of a 300 bp non-conserved sequence from the 5′ end of the 1.7 kb enhancer did not affect enhancer activity (Fig. 5, Tg 15). The data combining cross-species sequence comparison of *Math1*-flanking DNA, with the functional test of putative regulatory sequence in the transgenic mouse assay, have converged to define two conserved enhancers governing multiple domains of *Math1* expression.

Redundant activities of enhancer A and enhancer B govern expression in multiple *Math1* domains

The cross species sequence comparison identified the two blocks of homology that we refer to as enhancer A and enhancer B, and a construct containing all of enhancer A and half of enhancer B has sufficient information to drive *Math1*-specific expression at E10.5 (Fig. 5A, Tg 15). To continue to delineate functional regulatory sequences within the regions of conservation, we tested progressive deletions from the 3′ end of the 1.7 kb enhancer (Fig. 5A). A deletion from nucleotide 1193 to 1241 in the enhancer resulted in a consistent, dramatic loss of transgene expression (Fig. 5A). Low-level expression was detected in the hindbrain region in a small percentage of embryos. Deletion of the entire enhancer B sequence did not result in any additional loss of enhancer activity (Fig. 5A, Tg 13; Fig. 5F). Furthermore, a construct containing four tandemly arrayed copies of enhancer A alone gave hindbrain-only expression and no spinal neural tube expression (Fig. 5A, Tg 14; Fig. 5F). Having four copies of the enhancer increased the level of expression in the hindbrain and appears to increase the efficiency of getting expression (note 3/3 for Tg 14 versus 2/11 for Tg 13). To determine if enhancer A also functioned for other domains of *Math1* expression, we assayed enhancer A activity without enhancer B at E14.5 and P0. Transgene 13 not only functions at E10.5 for hindbrain expression, but it also is sufficient to drive reporter gene expression to the sensory epithelium of the ear and the Merkel cells in the whisker vibrissae at E14.5, and the EGL of the cerebellum at P0 (Fig. 5A).

The idea that the functions of the different subenhancers

within the 1.7 kb enhancer were distinct was suggested by enhancer A's ability to direct expression to multiple domains of *Math1* expression combined with its inability to direct expression in the spinal neural tube. To address this possibility, we tested the activity of enhancer B in the absence of enhancer A. A construct containing four tandemly arrayed copies of the functional portion of enhancer B was tested for expression at E10.5, E14.5 and P0. While only one out of nine embryos that expressed the transgene at E10.5 was identified, the expression was strong in the hindbrain and continued caudally down the spinal neural tube to the tail (Fig. 5A, Tg 17; Fig. 5G). In addition, this transgene was expressed in the ear and whisker vibrissae at E14.5, and EGL of the cerebellum at P0 (Fig. 5A). Although the transgene with four copies of enhancer B directed *lacZ* expression to all *Math1*-specific expression domains, ectopic expression of *lacZ* was consistently noted. This may reflect the absence of negative regulatory sequences in this transgene. Taken together, these data demonstrate that enhancer B is sufficient to direct expression of the *lacZ* reporter gene to the distinct domains of *Math1* expression, and most of this activity is redundant with the activity of enhancer A. There is no obvious sequence identity between the two enhancers that explain the redundant activity. Short sequences with related homeodomain consensus binding sites are present in both enhancers but the relevance of these sites has not been tested.

A bHLH consensus binding site in enhancer B is required for spinal neural tube expression at E10.5

Data described above demonstrate that a component necessary for enhancer activity at E10.5 must reside within the 49 bp deleted from nucleotides 1193 and 1241. This 49 bp sequence appears to be all that is necessary for the contribution of enhancer B to the overall enhancer activity at E10.5 (Fig. 5A,B, compare Tg 10, 11 and 13). Examination of the sequence within this 49 bp essential element identifies known transcription factor binding sites including an E-box (Murre et al., 1989), an N-box (Akazawa et al., 1992) and a hairy preferred site (van Doren et al., 1994) (Fig. 5B). Since N-boxes and hairy preferred sites are both recognized by putative negative regulators of transcription, we focused on the E-box, the consensus binding site for bHLH transcription factors. In an attempt to disrupt only the E-box site, we mutated the first three base pairs of the E-box, effectively abolishing the core sequence necessary for bHLH binding, while maintaining the integrity of the consensus binding site of the overlapping N-box (Fig. 5B). Expression from this transgene (Tg 12), whose sequence is identical to the fully active Tg 10 except for the 3 bp change in the E-box site, was dramatically reduced and mimicked the expression in transgenes lacking the entire enhancer B (Fig. 5A, Tg 12; Fig. 5E). These data demonstrate that all activity of enhancer B detected at E10.5 in this assay requires an intact E-box.

MATH1 binds the essential E-box site

The previous section addressed the requirement of the E-box site for activity of the enhancer, particularly in dorsal spinal neural tube. Since MATH1 itself is a bHLH transcription factor and has been shown to bind E-box consensus sites (Akazawa et al., 1995), we tested the ability of MATH1 to bind the E-box present in the enhancer using the electrophoretic mobility shift assay (EMSA). MATH1 and a putative binding partner E12 (Akazawa et al., 1995; Murre et al., 1989) were in vitro transcribed and translated using a rabbit reticulocyte lysate. The EMSA with labeled oligonucleotide identical to the E-box and its surrounding sequence in the enhancer (Fig. 5B) demonstrates the ability of MATH1/E12 heterodimers to bind this sequence (Fig. 6, lane 4). E12 and MATH1 homodimers also gel shift the E-box probe to varying degrees (Fig. 6, lanes 2, 3), but the MATH1/E12 heterodimer/E-box protein-DNA complex appears to be the most favorable interaction. The specificity and requirement of the E-box consensus within the oligonucleotide was demonstrated by competition with cold wild-type oligonucleotide, and the absence of competition with cold oligonucleotide in which three nucleotides of the E-box site were mutated (Fig. 6, compare lanes 4-6). The mutation incorporated into the oligonucleotide was the same as that tested in the transgenic mouse assay, Tg 12, and shown in Fig. 5B. Within the 1.7 kb *Math1* enhancer, there is a second E-box that is conserved between human and mouse (CATATG). It is located in enhancer A and deletion of sequence containing this E-box had no obvious consequence on enhancer activity in the transgenic assay (Fig. 5A, Tg 16 and 17). Consistent with these in vivo data, an oligonucleotide containing this E-box did not compete with binding of the MATH1/E12 heterodimer to the essential E-box site in the EMSA (Fig. 6, lane 7).

MATH1 itself is required for transgene expression

The EMSA demonstrates the ability of MATH1 to bind the essential E-box site in the *Math1* enhancer in vitro. To address the requirement for MATH1 function in the activity of the *Math1* regulatory sequences characterized here, we examined the expression of a *Math1/lacZ* transgene in the *Math1* mutant background. A line of *Math1/lacZ* transgenic mice containing Tg 6 (J2G) was crossed with the *Math1* mutant strain (Ben-Arie et al., 1997). Animals hemizygous for the *Math1/lacZ* transgenic allele and heterozygous for the *Math1* mutant allele were crossed with *Math1* heterozygous animals. Embryos were harvested at E10.5/E11.5, genotyped by PCR or Southern and scored for *lacZ* expression in the dorsal neural tube (Table 1). In the wild-type background, 100% of the embryos express the *lacZ* transgene in the dorsal neural tube (n=9). In contrast, none of the transgenic embryos had detectable β-gal activity in the *Math1* null background (n=8). Curiously, in the *Math1* heterozygous background, 20% percent of the embryos lacked transgene expression (n=25). The other 80% expressed the transgene at levels indistinguishable from wild type. This phenotype in the heterozygotes may reflect a *Math1* gene dosage effect in the

Table 1. Requirement for MATH1 function for *Math1* enhancer activity

*Math1*allele*	*lacZ*$^+$ embryos‡	β-gal$^+$ embryos§
+/+	9	9 (100%)
+/–	25	20 (80%)
–/–	8	0 (0%)

**Math1* null mutant strain (Ben-Arie et al., 1997)
‡J2G (Tg 6) *Math1/lacZ* transgenic allele
§Whole-mount β-gal assay revealing transgene expression in a *Math1*-specific pattern at E10.5

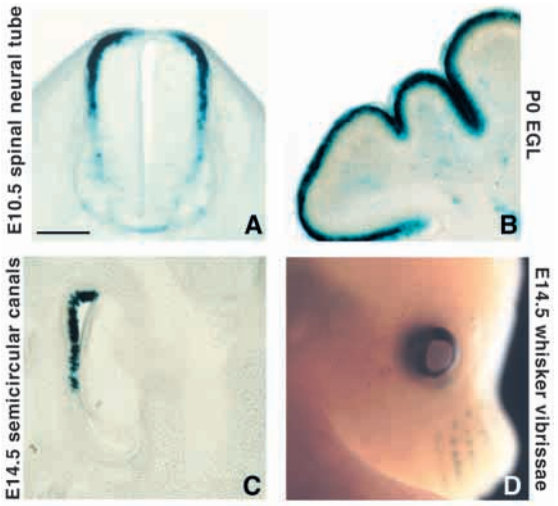

Fig. 4. The 1.7kb enhancer directs reporter gene expression to the diverse *Math1* domains. A transgenic founder male (strain J2Q-5) carrying Tg 9 was crossed with BDF1 females and transgenic progeny examined at E10.5 (A), P0 (B), and E14.5 (C,D) for β-gal activity. (A) 150 μm transverse vibratome section of the spinal neural tube showing *lacZ* expression in dorsal commissural interneurons. (B) 150 μm parasagittal vibratome section through the developing cerebellum showing *lacZ* expression restricted to the EGL. (C) 30 μm transverse cryosection through a semicircular canal ampulla showing *lacZ* expression in the sensory epithelium in the developing ear. (D) Whole-mount E14.5 embryo showing expression in the whisker vibrissae. Scale bar 85 μm (A), 70 μm (B), 80 μm (C) and 1.4 mm (D).

mixed strains used in these experiments. These data support a role for MATH1 in autoregulation.

DISCUSSION

A subset of neural-specific bHLH transcription factors including *Math1*, *Mash1*, *ngn1* and *ngn2* are expressed in proliferating neural progenitor cells and, as the cells progress to postmitotic differentiating neurons, expression of these bHLH factors is downregulated. Expression of this set of factors is mostly non-overlapping and their combined expression accounts for a majority of neural epithelium in the developing nervous system. Given the lethal phenotypes that occur in the absence of function of each of these factors and the lethality observed when *Math1* is misexpressed throughout the neuroepithelium (Isaka et al., 1999), precise control is critical for expression of these genes. Identification of the mechanisms of transcriptional regulation of neural regulatory genes like *Math1* is a first step toward identifying the signaling pathways that control early neuronal development. We have identified two sequences that are conserved between human and mouse and are sufficient for driving expression of a reporter gene to the diverse *Math1* expression domains. A bHLH factor consensus binding site, the E-box, is essential for transgene expression in the spinal neural tube and appears to reflect a role for autoregulation in control of *Math1* expression. Regulatory sequence driving expression in the other *Math1* domains, including the EGL of the cerebellum and the sensory epithelium in the developing ear and whisker vibrissae, is redundant in the two conserved sequences and does not require the E-box site.

Math1 expression in a mechanosensory cell in whisker vibrissae and hairy skin

Of the vertebrate bHLH transcription factors, MATH1 is the most closely related to the *Drosophila* proneural factor ATONAL. *atonal* is expressed in the proneural clusters of the chordotonal (mechanosensory) organs, as well as in the developing compound eye (Jarman et al., 1993, 1994, 1995). *Math1* expression in neural precursors destined to become dorsal commissural interneurons and in the precursors of the granular layer of the cerebellum (Akazawa et al., 1995; Ben-Arie et al., 1997; Helms and Johnson, 1998) does not have obvious correlates with the domains of *atonal* expression. More recently, it has been reported that *Math1* is also expressed in the developing hair cells of the vestibular and auditory systems (Bermingham et al., 1999). This expression in sensory epithelium is more analogous to the cell-type specificity of the *Drosophila* gene.

Here, we present yet another domain of *Math1* expression not previously reported, which appears to be more analogous to *atonal*-expressing cell types. Epidermal Merkel cells are the mechanosensory cells of the whisker vibrissae and hairy skin that are thought to transmit tactile sensations in vertebrates (Gu et al., 1981; Hartschuh et al., 1983). We demonstrate that MATH1 protein is expressed early in these cells in the developing whisker vibrissae and hairy skin. Little is known about the development of this cell type. Recently, a role for MATH1 in Merkel cell development has been demonstrated (N. B. A. and H. Y. Z., unpublished data). MATH1 will provide an important tool to probe the development of these mechanosensory cells. The expression of *Math1* in the Merkel cells may share the most direct similarity to the expression of *atonal* in *Drosophila* and, thus, like ATONAL may serve a proneural function for this cell type. Comparison of the *Math1* enhancer sequence to sequences identified in *atonal* that regulate expression in chordotonal organs may reveal conserved upstream control mechanisms.

Math1 enhancer activity in the spinal neural tube is lacking negative regulatory information

Math1 regulatory sequences flanking the coding region drive expression of a *lacZ* reporter gene to the *Math1* expression domain in the dorsal neural tube. However, this expression persists into more differentiated neurons derived from the *Math1*-expressing progenitors (Helms and Johnson, 1998). This discrepancy in *Math1* versus *lacZ* reporter expression in the transgenic embryos has raised the question of whether important negative regulatory elements are missing from the transgene, or whether the persistence in reporter gene expression is simply due to differences in stability of *lacZ* mRNA or β-gal protein. To distinguish between these two possibilities, we used an epitope-tagged *Math1* reporter gene in place of the *lacZ* reporter. The epitope-tagged *Math1* provided a reporter with stability at the mRNA and protein level that most closely mimics the stability of the endogenous gene. The results from this experiment are consistent with the aberrant pattern of expression being due to transgene regulation, and most likely reflect missing negative regulatory information in the transgene. This missing regulatory

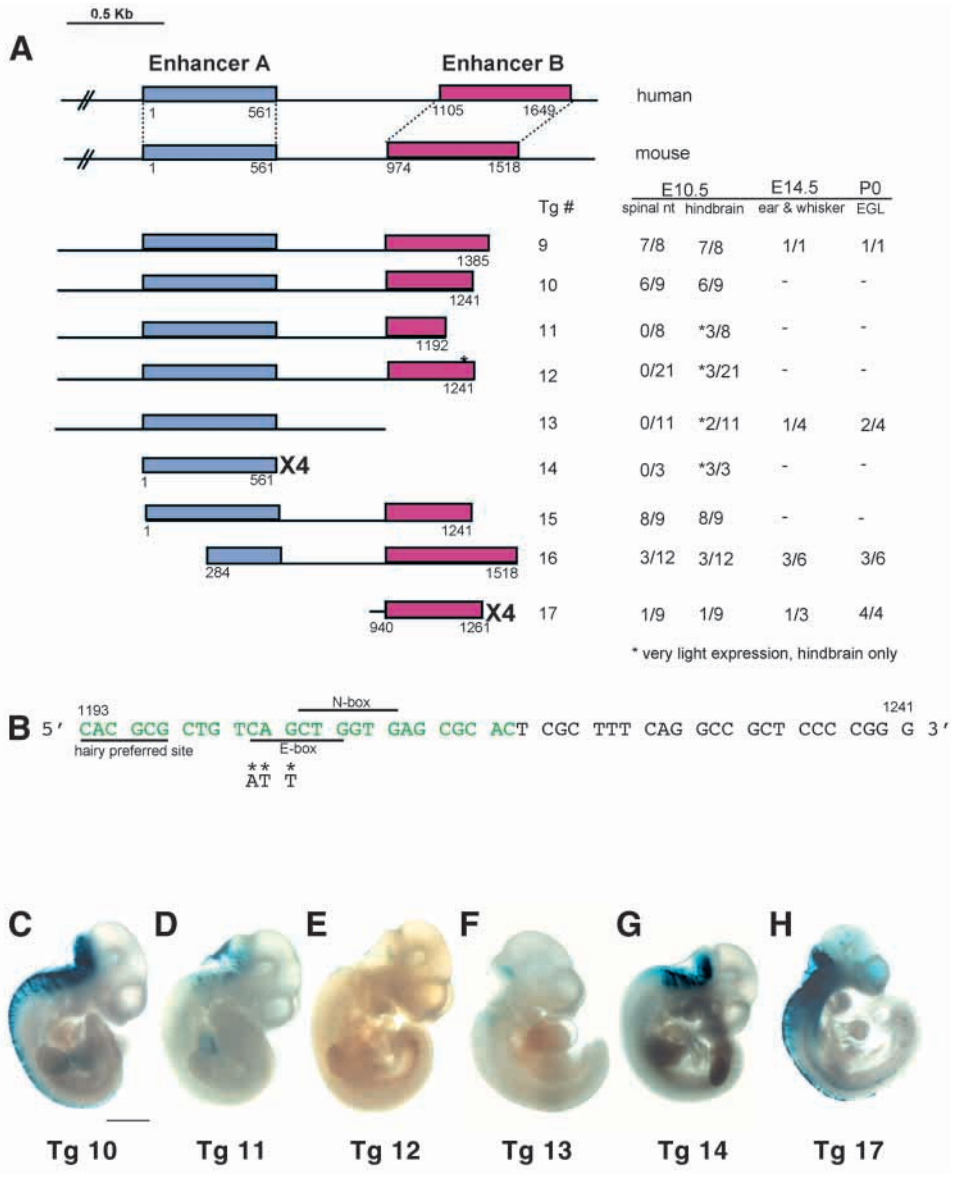

Fig. 5. Identification of two conserved enhancer elements and an essential E-box regulate *Math1* expression. (A) Diagram comparing the conserved regions between mouse and human sequences located ~3.4 kb 3′ of the site encoding the translation stop for *Math1* or *Hath1*. Conserved sequence blocks are defined as enhancer A (blue) and enhancer B (red). The *Math1* sequences tested in transgenes 9-17 were fused to the *β-globin/lacZ* reporter (BGZA). Most embryos were assayed as founders (G_0). Transgenic lines were made from Tg 9, 13 and 16 for analysis of expression at multiple developmental stages. Transgenic data are presented as the number of embryos expressing β-gal activity in a *Math1*-specific pattern per the number of transgenic embryos that were identified by DNA analysis. The different stages and cell types examined are shown. (B) Sequence of the region deleted between Tg 10 and 11 from nucleotides 1193 to 1241. The location of the hairy preferred site (Van Doren et al., 1994), E-box (Murre et al., 1989) and N-box (Akazawa et al., 1992) is indicated. The three nucleotides mutated in the E-box in Tg 12 are shown below the sequence. This mutated the E-box CAGCTG to ATTCTG, which maintained the integrity of the overlapping N-box. The sequence in green is the E-box oligonucleotide used in the EMSA in Fig. 6. (C-H) E10.5 whole-mount β-gal-stained embryos containing transgenes 10 (C), 11 (D), 12 (E), 13 (F), 14 (G) and 17 (H). Deletion of 49 bp at the 3′ end of Tg 10 completely abolished transgene expression in the spinal neural tube and resulted in reduced expression in the hindbrain (compare C and D). (E) Mutation of three base pairs in the E-box consensus site within the 49 bp region resulted in the same loss of enhancer activity. (G) Enhancer A was tested in four copies and it drives expression only in the hindbrain at E10.5 but not the spinal neural tube. (H) Enhancer B in four copies drives strong expression in both the hindbrain and spinal neural tube. Note the appearance of ectopic expression in the ventral neural tube. Scale bar 1.2 mm.

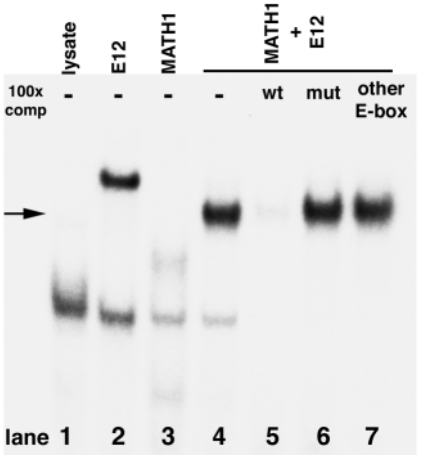

Fig. 6. MATH1/E12 heterodimers bind the essential E-box in the *Math1* enhancer. Electrophoretic mobility shift assay with reticulocyte lysate alone (lane 1), with E12 (lane 2), with MATH1 (lane 3) or with MATH1 plus E12 (lanes 4-7) is shown. The probe is a 25 bp oligonucleotide containing the E-box and its flanking sequence shown in green in Fig. 5B. 100-fold molar excess of cold wild-type oligonucleotide (lane 5), cold oligonucleotide mutant in three nucleotides of the E-box site (lane 6), or a cold oligonucleotide containing an E-box site also present in the *Math1* enhancer but lacks function (lane 7), were added to the MATH1/E12-binding reaction to evaluate the specificity of the protein/DNA interaction. The arrow indicates the MATH1/E12/E-box complex. The unbound probe is not shown.

information is not present within 26 kb of *Math1*-flanking sequence (Tg 1, Fig. 3A and data not shown). Additionally, this aberrant regulation of the transgene in the neural tube seems to be limited to that domain of expression, as *lacZ* expression in the other *Math1* domains appears to correspond precisely with *Math1* expression. We do not understand what type of negative regulatory information may be missing. It is curious that this persistence in *lacZ* reporter expression into differentiated neurons in the neural tube has also been seen in regulation studies with other bHLH transcription factors including *ngn1* and *ngn2* (J. E. J., unpublished data).

Sequence homologies reveal conserved enhancers directing *Math1* expression

By cloning and sequencing the flanking sequences of *Hath1*, the human ortholog of *Math1*, we were able to identify two highly conserved regions in the 3′ flanking sequence that represent the regulatory regions that we identified as being important for *Math1* expression. Besides the high degree of sequence conservation (92% and 87%, respectively, for enhancers A and B), the location 3′ of the coding region (~3.4 kb from the coding sequence) was also conserved. It is striking that the degree of similarity between the enhancers in the two species is greater than the degree of similarity between the coding sequences (85%). Also notable is the relatively large size of the conserved regions, 561 and 544 bp for enhancer A and B, respectively. We expected to find much smaller regions of similarity that would lead to potential regulatory factor binding sites as has been done previously in comparing sequences from *Drosophila melanogaster* and *Drosophila virilis* (Cripps et al., 1998). The conservation of such large regions of sequence suggests that multiple transcription factor binding sites are present and may be important in combination to provide the precise temporal and spatial information for *Math1* expression.

The presence of two conserved blocks of sequence in the *Math1* enhancer initially suggested the possibility that each block was important for a specific subset of the *Math1* expression pattern, and thus represented the modular nature of the enhancer. One of the main goals of identifying *cis*-acting regulatory sequences for genes with complex expression patterns is the idea that the *cis*-elements for different domains may be separable from each other. Identifying the "regulatory module" for each expression domain would be useful for understanding mechanisms of regulation and provide an excellent tool for future studies focused on specific domains of expression. For the handful of control genes whose regulatory regions have been analyzed, "regulatory modules" for different domains of expression have been identified (Black and Olson, 1998; MacKenzie et al., 1997; Manzanares et al., 1999; Sun et al., 1998; Zhang et al., 1997). In contrast, our analysis of *Math1* regulation has revealed little if any modularity of *cis*-elements for different expression domains. Each enhancer (A or B) alone is sufficient to drive expression in the different *Math1* expression domains independent of the other enhancer. The exception is the expression in spinal neural tube, which requires the E-box in enhancer B. This redundant function is not due to any significant sequence homology between the two enhancers. The presence of redundant enhancers that show no apparent sequence homology has also been seen in the enhancer characterized for the *Wnt1* gene (Rowitch et al., 1998). Some core binding sequences for homeodomain transcription factors are found in both enhancers A and B, and similar proteins may bind them. The relevance of these sites and the identification of factors that bind them have not been established. If, in fact, the elements regulating *Math1* expression are not resolved into distinct modules, it may be that there is a common mechanism for regulation in these disparate domains of *Math1* expression. Indeed, there are many transcription factors (MSX and PAX families) and signaling pathways (BMP, Wnt, FGF) that are common to all of the *Math1* expression domains (Engelkamp, 1999; Goulding et al., 1991; Hogan, 1996; Lee and Jessell, 1999; Maruouka et al., 1998; McMahon, 1992; Wang et al., 1996).

Identification of an essential E-box binding site and autoregulation of MATH1

Expression of *Math1* in the spinal neural tube is the only domain of expression that appears to be separable from all other domains. Expression in this domain requires an intact E-box consensus site. When this E-box site, located in enhancer B, is specifically mutated, expression of the *lacZ* transgene is abolished in the spinal neural tube, and dramatically reduced in the hindbrain, at E10.5. The mutation of this E-box site has the same effect as deleting the entire enhancer B (compare Tg 12 and Tg 13), and thus it appears as though the detectable activity of enhancer B requires this E-box. This E-box appears to be unnecessary for expression in the other domains of *Math1* expression (see Tg 13). It may, however, play a role in modulating levels of expression in other domains since all constructs missing enhancer B have somewhat reduced levels of expression, though by no means as drastic as the reduction in spinal neural tube expression at E10.5. It is possible that regulation of reporter gene expression through the E-box may reduce the overall level of expression at E10.5 resulting in a decrease in transgene expression in the hindbrain, and complete loss of expression in the more caudal neural tube.

It has been well established that the E-box is the consensus binding site for the bHLH family of transcription factors (Murre et al., 1989). Other bHLH factors expressed in the dorsal neural tube at E10.5 include *ngn1* and *ngn2*, two other members of the *atonal*-related class of vertebrate bHLH genes (Gradwohl et al., 1996; Ma et al., 1996; Sommer et al., 1996). Although present in the same region of the spinal neural tube, a detailed analysis suggests that the expression domains do not significantly overlap (K. Gowan and J. E. J., unpublished data) and, thus, there are no known neural-specific bHLH factors co-expressed in the *Math1* domain. This leaves MATH1 as the most obvious candidate for regulation through the E-box site. The idea that MATH1 may autoregulate is not surprising, since its *Drosophila* counterpart, ATONAL, reportedly autoregulates expression in multiple domains (Sun et al., 1998). Furthermore, in *Xenopus*, embryos ectopically expressing *Math1* induce the expression of endogenous *Xath1* (Kim et al., 1997).

Here we demonstrate the requirement for MATH1 in the activity of the identified *Math1* enhancer, consistent with a role for MATH1 in autoregulation. The complete loss of transgene expression in the mutant background supports a role for autoregulation in enhancer B as predicted from the identification of the essential E-box in this enhancer. However, it suggests enhancer A has an autoregulatory component for the hindbrain expression. If so, this autoregulatory component

only functions in the hindbrain, and not in the spinal cord, as seen in the strong hindbrain expression when four copies of enhancer A are present (Tg 14). Although the complete loss of *lacZ* expression could result if the relevant population of cells fail to form in the absence of MATH1, no obvious phenotype in the dorsal spinal neural tube has been reported in the *Math1* mutant strain (Ben-Arie et al., 1997). Support for the maintenance of the progenitor population in the *Math1* mutant comes from analysis of the mouse mutant with *lacZ* knocked into the *Math1* locus (Bermingham et al., 1999). In this mutant, *lacZ* expression is retained in a population of cells in the dorsal neural tube (N. B. A. and H. Y. Z., unpublished observations). The continued expression of the *lacZ* in the knock-in suggests that the endogenous locus must contain elements that are independent of MATH1 and are not found in the *Math1/lacZ* transgenes tested here.

The responsiveness of the *Math1/LacZ* transgene to regulation by MATH1 itself brings up the possibility that the enhancer activity that we have identified does not have the ability to initiate expression but rather is important only for maintenance or high levels of expression. This idea is supported by the complete loss of expression of the transgene in the *Math1* mutant background. In contrast to this suggestion, however, the region of homology between mouse and human is much more extensive than the small region around the E-box. This suggests additional factors may be interacting at the enhancer to modify *Math1* expression. Taken together, the data suggest that the 1.7 kb enhancer, containing large blocks of sequence conservation between human and mouse, is important for expression in multiple cell- and stage-specific regions in the developing nervous system. MATH1 binding to the E-box in enhancer B is required for high level expression in all domains, particularly for expression in spinal neural tube. The enhancer region identified in this study is sufficient to direct expression to apparently diverse regions of the developing nervous system suggesting there may be common mechanisms for regulating development of these tissues. In addition, this work serves as a foundation for further studies defining the upstream mechanisms controlling the precise spatial and temporal expression of this essential regulator of nervous system development.

We gratefully acknowledge Drs L. Parada, E. Olson and D. Rowitch for critical comments on the manuscript, members of the Johnson laboratory for valuable discussions of this research, T. Savage and K. Gowan for technical assistance, S. Hall in the Sequencing Facility at UTSW for providing quality sequence data, Dr E. Lumpkin for valuable advice and protocols for Merkel cell analysis, Dr D. Srivastava for the human genomic library, Dr P. Rigby for the BGZA *lacZ* reporter construct, and Dr H. Krämer for the Bride of Sevenless epitope and the antibodies that recognize it. This work was supported by a predoctoral NIH training grant #T32GM07062 to A. W. H., a grant to J. E. J. from the Muscular Dystrophy Association. J. E. J. is an Established Investigator of the American Heart Association.

REFERENCES

Airaksinen, M. S., Koltzenburg, M., Lewin, G. R., Masu, Y., Helbig, C., Wolf, E., Brem, G., Toyka, K. V., Thoenen, H. and Meyer, M. (1996). Specific subtypes of cutaneous mechanoreceptors require Neurotrophin-3 following peripheral target innervation. *Neuron* **16**, 287-295.

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

Akazawa, C., Sasai, Y., Nakanishi, S. and Kageyama, R. (1992). Molecular characterization of a rat negative regulator with basic-helix-loop-helix structure predominantly expressed in the developing nervous system. *J. Biol. Chem.* **267**, 21879-21885.

Ben-Arie, N., Bellen, H. J., Armstrong, D. L., McCall, A. E., Gordadze, P. R., Guo, Q., Matzuk, M. M. and Zoghbi, H. Y. (1997). *Math1* is essential for genesis of cerebellar granule neurons. *Nature* **390**, 169-172.

Ben-Arie, N., McCall, A. E., Berkman, S., Eichele, G., Bellen, H. J. and Zoghbi, H. Y. (1996). Evolutionary conservation of sequence and expression of the bHLH protein Atonal suggests a conserved role in neurogenesis. *Human Molecular Genetics* **5**, 1207-1216.

Bermingham, N. A., Hassan, B. A., Price, S. D., Vollrath, M. A., Ben-Arie, M., Eatock, R. A., Bellen, H. J., Lysakowski, A. and Zoghbi, H. Y. (1999). Math1: An essential gene for the generation of inner ear hair cells. *Science* **284**, 1837-1841.

Black, B. L., Molkentin, J. D. and Olson, E. N. (1998). Multiple roles for the MyoD basic region in transmission of transcriptional activation signals and interation with MEF2. *Molec. Cell. Biol.* **18**, 69-77.

Black, B. L. and Olson, E. N. (1998). Transcriptional control of muscle development by myocyte enhancer factor-2 (MEF2) proteins. *Ann. Rev. Cell Dev. Biology* **14**, 167-196.

Cabrera, C. V. and Alonso, M. C. (1991). Transcriptional activation by heterodimers of the *achaete-scute* and *daughterless* gene products of *Drosophila*. *EMBO J.* **10**, 2965-2973.

Campos-Ortega, J. A. and Jan, Y. N. (1991). Genetic and molecular bases of neurogenesis in *Drosophila melanogaster*. *Annu. Rev. Neurosci.* **14**, 399-420.

Casarosa, S., Fode, C. and Guillemot, F. (1999). *Mash1* regulates neurogenesis in the ventral telencephalon. *Development* **126**, 525-534.

Cripps, R. M., Black, B. L., Zhao, B., Lien, C. L., Schulz, R. A. and Olson, E. N. (1998). The myogenic regulatory gene Mef2 is a direct target for transcriptional activation by Twist during Drosophila myogenesis. *Genes Dev.* **12**, 422-434.

Cubadda, Y., Heitzler, P., Ray, R. P., Bourouis, M., Ramain, P., Gelbart, W., Simpson, P. and Haenlin, M. (1997). *u-shaped* encodes a zinc finger protein that regulates the proneural genes *achaete* and *scute* during the formation of bristles in *Drosophila*. *Genes Dev.* **11**, 3083-3095.

Engelkamp, D., Rashbass, P, Seawright, A, van Heyningen, V. (1999). Role of Pax 6 in development of the cerebellar system. *Development* **126**, 3585-3596.

Fode, C., Gradwohl, G., Morin, X., Dierich, A., LeMeur, M., Goridis, C. and Guillemot, F. (1998). The bHLH protein NEUROGENIN2 is a detemination factor for epibranchial placode-derived sensory neurons. *Neuron* **120**, 483-494.

Gomez-Skarmeta, J. L., del Corral, R. D., de la Calle-Mustienes, E., Ferre-Marco, D. and Modolell, J. (1996). Araucan and caupolican, two members of the novel iroquois complex, encode homeoproteins that control proneural and vein-forming genes. *Cell* **85**, 95-105.

Gomez-Skarmeta, J. L., Rodriguez, I., Martinez, C., Culi, J., Ferres-Marco, D., Beamonte, D. and Modolell, J. (1995). Cis-regulation of achaete and scute: shared enhancer-like elements drive their co-expression in proneural clusters of the imaginal discs. *Genes Dev.* **9**, 1869-1882.

Goulding, M. D., Chalepakis, G., Deutsch, U., Ersehius, J. R. and Gruss, P. (1991). Pax3, a novel murine DNA binding protein expressed during early neurogenesis. *EMBO J.* **10**, 1135-1147.

Gradwohl, G., Fode, C. and Guillemot, F. (1996). Restricted Expression of a novel murine *atonal*-related bHLH protein in undifferentiated neural precursors. *Dev. Biol.* **180**, 227-241.

Gu, J., Polak, J. M., Tapia, F. J., Marangos, P. J. and Pearse, A. G. E. (1981). Neuron specific enolase in the Merkel cells of mammalian skin. *Am. J. Pathology* **104**, 63-68.

Guillemot, F., Lo, L. C., Johnson, J. E., Auerbach, A., Anderson, D. J. and Joyner, A. L. (1993). Mammalian achaete-scute homolog 1 is required for the early development of olfactory and autonomic neurons. *Cell* **75**, 463-476.

Hartschuh, W., Weihe, E., Yanaihara, N. and Reinecke, M. (1983). Immunohistochemical localization of vasoactive intestinal polypeptide (VIP) in Merkel cells of various mammals: evidence for a neuromodulatory function of the Merkel cell. *J. Invest. Dermatol.* **81**, 361-364.

Helms, A. W. and Johnson, J. E. (1998). Progenitors of dorsal commissural

interneurons are defined by MATH1 expression. *Development* **125**, 919-925.

Henrique, D., Tyler, D., Kintner, C., Heath, J. K., Lewis, J. H., Ish-Horowicz, D. and Storey, K. G. (1997). cash4, a novel *achaete-scute* homolog induced by Hensen's node during generation of the posterior nervous system. *Genes Dev.* **11**, 603-615.

Hogan, B. (1996). Bone Morphogenetic Proteins: multifunctional regulators of vertebrate development. *Genes Dev.* **10**, 1580-1594.

Hogan, B., Costantini, F. and Lacy, E. (1986). *Manipulating the Mouse Embryo: A Laboratory Manual*. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.

Horton, S., Meredith, A., Richardson, J. A. and Johnson, J. E. (1999). Correct coordination of neuronal differentiation events in ventral forebrain requires the bHLH factor MASH1. *Molec. Cell. Neurosci.* **14**, 355-369.

Isaka, F., Ishibashi, M., Taki, W., Hashimoto, N., Nakanishi, S. and Kageyama, R. (1999). Ectopic expression of the bHLH gene Math1 disturbs neural development. *Eur. J. Neurosci.* **11**, 2582-2588.

Jan, L. and Jan, Y. N. (1993). HLH proteins, fly neurogenesis, and vertebrate myogenesis. *Cell* **75**, 827-830.

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.

Jarman, A. P., Grell, E. H., Ackerman, L., Jan, L. Y. and Jan, Y. N. (1994). *atonal* is the proneural gene for *Drosophila* photoreceptors. *Nature* **369**, 398-400.

Jarman, A. P., Sun, Y., Jan, L. Y. and Jan, Y. N. (1995). Role of the proneural gene, *atonal*, in formation of *Drosophila* chordotonal organs and photoreceptors. *Development* **121**, 2019-2030.

Johnson, J. E., Birren, S. J. and Anderson, D. J. (1990). Two rat homologues of *Drosophila achaete-scute* specifically expressed in neuronal precursors. *Nature* **346**, 858-861.

Kageyama, R., Sasai, Y., Akazawa, C., Ishibashi, M., Takebayashi, K., Shimizu, C., Tomita, K. and Nakanishi, S. (1995). Regulation of mammalian neural development by helix-loop-helix transcription factors. *Critical Reviews in Neurobiology* **9**, 177-188.

Kim, P., Helms, A. W., Johnson, J. E. and Zimmerman, K. (1997). *XATH1*, a vertebrate homolog of *Drosophila atonal*, induces neuronal differentiation within ectodermal progenitors. *Dev. Biol.* **187**, 1-12.

Krämer, H., Cagan, R. L. and Zipursky, S. L. (1991). Interaction of *bride of sevenless* membrane-bound ligand and the sevenless tyrosine-kinase receptor. *Nature* **352**, 207-212.

Lee, J. E. (1997). Basic helix-loop-helix genes in neural development. *Curr. Opin. Neurobiol.* **7**, 13-20.

Lee, J. E., Hollenberg, S. M., Snider, L., Turner, D. L., Lipnick, N. and Weintraub, H. (1995). Conversion of Xenopus ectoderm into neurons by NeuroD, a basic-helix-loop-helix protein. *Science* **268**, 836-844.

Lee, K. J. and Jessell, T. M. (1999). The specification of dorsal cell fates in the vertebrate central nervous system. *Ann. Rev. Neuroscience* **22**, 261-294.

Leyns, L., Gomez-Skarmeta, J. L. and Dambly-Chaudiere, C. (1996). Iroquois: a prepattern gene that controls the formation of bristles on the thorax of Drosophila. *Mechan. Dev.* **59**, 63-72.

Ma, Q., Chen, Z., del Barco Barrantes, I., de la Pompa, J. L. and Anderson, D. J. (1998). *neurogenin1* is essential for the determination of neuronal precursors for proximal cranial sensory ganglia. *Neuron* **120**, 469-482.

Ma, Q., Kintner, C. and Anderson, D. J. (1996). Identification of *neurogenin*, a vertebrate neuronal determination gene. *Cell* **87**, 43-52.

MacKenzie, A., Purdie, L., Davidson, D., Collinson, M. and Hill, R. E. (1997). Two enhancer domains control early aspects of the complex expression pattern of Msx1. *Mech. Dev.* **62**, 29-40.

Manzanares, M., Cordes, S., Ariza-McNaughton, L., Sadl, V., Maruthainar, K., Barsh, G. and Krumlauf, R. (1999). Conserved and distinct roles of kriesler in regulation of the paralogous Hoxa3 and Hoxb3 genes. *Development* **126**, 759-769.

Martinez, C. and Modolell, J. (1991). Cross-regulatoy interactions between the proneural *achaete* and *acute* genes of *Drosophila*. *Science* **251**, 1485-1487.

Maruouka, Y., Ohbayashi, N., Hoshikawa, M., Itoh, N., Hogan, B. L. M. and Furuta, Y. (1998). Comparison of the expression of three highly related genes, Fgf8, Fgf17, and Fgf18 in the mouse embryo. *Mech. Dev.* **74**, 175-177.

McCormick, M. B., Tamini, R. M., Snider, L., Asakura, A., Bergstorm, D. and Tapscott, S. J. (1996). neuroD2 and neuroD3: distinct expression patterns and transcriptional activation potentials within the neuroD gene family. *Molec. Cell. Biol.* **16**, 5792-5800.

McMahon, A. P. (1992). The Wnt family of developmental regulators. *Trends in Genetics* **8**, 236-242.

Mercer, E. H., Hoyle, G. W., Kapur, R. P., Brinster, R. L. and Palmiter, R. D. (1991). The dopamine ß-hydroxylase gene promoter directs expression of E. coli *lacZ* to sympathetic and other neurons in transgenic mice. *Neuron* **7**, 703-716.

Murre, C., McCaw, P. S. and Baltimore, D. (1989). A new DNA binding and dimerization motif in immunoglobin enhancer binding, *daughterless*, *MyoD* and *myc* proteins. *Cell* **56**, 777-783.

Naya, F. J., Wu, C., Richardson, J. A., Overbeek, P. and Olson, E. N. (1999). Transcriptional activity of MEF2 during mouse embryogenesis monitored with a MEF-2 dependent transgene. *Development* **126**, 2045-2052.

Ramain, P., Heitzler, P., Haenlin, M. and Simpson, P. (1993). *pannier*, a negative regulatory of *achaete* and *scute* in *Drosophila*, encodes a zinc finger protein with homology to the vertebrate transcription factor GATA-1. *Development* **119**, 1277-1291.

Rowitch, D. H., Echelard, Y., Danielian, P. S., Gellner, K., Brenner, S. and McMahon, A. P. (1998). Identification of an evolutionarily conserved 110 base-pair *cis*-acting regulatory sequence that governs *Wnt-1* expression in the murine neural plate. *Development* **125**, 2735-2746.

Shimizu, C., Akazawa, C., Nakanishi, S. and Kageyama, R. (1995). MATH-2, a mammalian helix-loop-helix factor structurally related to the product of the *Drosophila* proneural gene *atonal*, is specifically expressed in the nervous system. *Eur. J. Biochem.* **229**, 239-248.

Skeath, J. B. and Carroll, S. B. (1991). Regulation of achaete-scute gene expression and sensory organ pattern formation in the Drosophila wing. *Genes and Development* **5**, 984-995.

Sommer, L., Ma, Q. and Anderson, D. J. (1996). neurogenins, a novel family of atonal-related bHLH transcription factors, are putative mammalian neuronal determination genes that reveal progenitor cell heterogenity in the developing CNS and PNS. *Molec. Cell. Neurosci.* **8**, 221-241.

Sun, Y., Jan, L. Y. and Jan, Y. N. (1998). Transcriptional regulation of *atonal* during development of the *Drosophila* peripheral nervous system. *Development* **125**, 3731-3740.

Torii, M., Matsuzaki, F., Osumi, N., Kaibuchi, K., Nakamura, S., Casarosa, S., Guillemot, F. and Nakafuku, M. (1999). Transcription factors Mash-1 and Prox-1 delineate early steps in differentiation of neural stem cells in the developing central nervous system. *Development* **126**, 443-456.

Tuttle, R., Nakagawa, Y., Johnson, J. E. and O'Leary, D. D. M. (1999). Defects in thalamocortical axon pathfinding correlate with altered cell domains in embryonic *MASH1* deficient mice. *Development* **126**, 1903-1916.

van Doren, M., Bailey, A. M., Esnayra, J., Ede, K. and Posakony, J. W. (1994). Negative regulation of proneural gene activity: hairy is a direct transcriptional repressor of *achaete*. *Genes Dev.* **8**, 2729-2742.

van Doren, M., Powell, P. A., Pasternak, D., Singson, A. and Posakony, J. W. (1992). Spatial regulation of proneural gene activity: auto- and cross-activation of achaete is antagonized by extramacrochaetae. *Genes Dev.* **6**, 2592-2605.

Verma-Kurvari, S., Savage, T., Gowan, K. and Johnson, J. E. (1996). Lineage-specific regulation of the neural differentiation gene MASH1. *Dev. Biol.* **180**, 605-617.

Verma-Kurvari, S., Savage, T., Smith, D. and Johnson, J. E. (1998). Multiple elements regulate *Mash1* expression in the developing CNS. *Dev. Biol.* **197**, 106-116.

Wang, W., Chen, X., Xu, H. and Lufkin, T. (1996). Msx3: a novel murine homolog of the Drosophila msh homeobox gene restricted to the dorsal embryonic central nervous system. *Mechan. Dev.* **58**, 203-215.

Yee, S. and Rigby, P. W. J. (1993). The regulation of myogenin gene expression during the embryonic development of the mouse. *Genes Dev.* **7**, 1277-1289.

Zhang, F., Popperl, H., Morrison, A., Kovacs, E. N., Prideaux, V., Schwarz, L., Krumlauf, R., Rossant, J. and Featherstone, M. S. (1997). Elements both 5′ and 3′ to the murine Hoxd4 gene establish anterior borders of expression in mesoderm and neurectoderm. *Mech. Dev.* **67**, 49-58.