Human Molecular Genetics, 1996, Vol. 5, No. 9 **1207–1216**

ARTICLE

Evolutionary conservation of sequence and expression of the bHLH protein Atonal suggests a conserved role in neurogenesis

Nissim Ben-Arie[1,2], Alanna E. McCall[1], Scott Berkman[3], Gregor Eichele[3], Hugo J. Bellen[2,4] and Huda Y. Zoghbi[1,2,4,*]

[1]Department of Pediatrics, [2]Department of Molecular and Human Genetics, [3]Department of Biochemistry and [4]Howard Hughes Medical Institute, Baylor College of Medicine, One Baylor Plaza, Houston, Texas 77030, USA

Received May 14, 1996; Revised and Accepted June 20, 1996

GenBank accession no. U61148–U61152

***atonal* is a *Drosophila* proneural gene that belongs to the family of basic helix-loop-helix (bHLH)- containing proteins. It is expressed in the chordotonal organs and photoreceptor cells, and flies that lack Atonal protein are ataxic and blind. Here we report the cloning of *atonal* homologs from red flour beetle, puffer fish, chicken, mouse, and human. The bHLH domain is conserved throughout evolution, while the entire coding region is highly similar in mammals. Both the chicken and the mouse homologs are expressed early in embryogenesis in the hind brain, and specifically in cells predicted to give rise to the external granular layer of the cerebellum. In addition, these genes are expressed throughout the dorsal part of the spinal cord, in patterns different from those found for other genes, like *LH-2* and *wnt-1*. The mouse homolog (*Math1*) maps to mouse chromosome 6, and the human homolog (*HATH1*) to human chromosome 4q22. Two neurological mouse mutants, *Lc* and *chp*, were found to map to the vicinity of *Math1*, but are not caused by mutations in *Math1*. The evolutionary conservation of this gene and its mRNA expression patterns during embryogenesis suggests that it plays a key role in the development of the vertebrate central nervous system.**

INTRODUCTION

The molecular mechanisms that control mammalian neural development involve many genes, most of which remain to be isolated and characterized (1). *Drosophila* and mammals seem to use similar molecular mechanisms to determine which cells will become part of the nervous system (2). Many of the genes required for the specification of neuronal identity (3), neuronal differentiation (4), and growth cone guidance (5) in *Drosophila* have vertebrate homologs, whose sites of expression and function are reminiscent of their *Drosophila* homologs (1,5–8). Hence, progress in understanding neural development in *Drosophila* provides an excellent basis for studies aimed at investigating the mechanisms that control neuronal development in the mammalian nervous system.

One class of genes that is of particular interest in neuronal development encodes a family of proteins that contain a basic helix-loop-helix (bHLH) motif (9–12). These proteins are transcription factors with a basic domain necessary for DNA binding, and two helices that allow the formation of heterodimers with other bHLH proteins.

atonal is a proneural gene that belongs to the family of bHLH-containing proteins, and that plays an essential role in the development of the *Drosophila* nervous system (13). In *Drosophila* embryos *atonal* is expressed in the ectodermal proneural clusters and sensory organ precursors that give rise to the chordotonal organs, which are receptors for stretch and/or vibration (proprioception), as well as in the optic furrow of the eye-antennal disc and in the inner proliferative zone of the developing brain lobe (13–15). Deletion of genomic region spanning the *atonal* locus causes a lack of a subset of the PNS organs that includes all the ventral chordotonal organs and some multiple dendritic neurons (13). In addition, these flies lack photoreceptors and develop an apoptotic atrophy of the imaginal disc (15). Adult flies that lack *atonal* are viable; however, they are blind, uncoordinated and fly poorly (15).

In this study we show that *Drosophila atonal* is evolutionarily conserved and that its sequence shares high similarity within the bHLH domain with homologs from *Tribolium castenium* (red flour beetle), *Fugu rubripes* (puffer fish), chicken, mouse, and human. Expression analysis revealed that the mouse and chicken homologs are expressed in the dorsal regions of the hind brain and

*To whom correspondence should be addressed

spinal cord. We have mapped the mouse *atonal* homolog (*Math1*) to the central portion of mouse chromosome 6 near two mouse mutants with neurological features: lurcher (*Lc*) and cerebellar-hypoplasia *(chp)*. To determine if the phenotype in *Lc/+* and *chp/chp* mutant mice is caused by a defect in *Math1*, the coding region from *Lc* and *chp* was sequenced and the genomic locus was studied by Southern analysis. No mutations nor genomic rearrangements were found in *Math1* for either *Lc* or *chp*, suggesting that these genes are not allelic with *Math1*. The human *atonal* homolog (*HATH1*) was cloned and mapped. It is highly similar to the mouse *Math1* gene (89% identity) and maps to human chromosome 4q22. The expression patterns of the vertebrate homologs of *Drosophila atonal* and the high conservation of the coding sequence within the bHLH domain suggest that this gene plays an important role in the development of the vertebrate nervous system.

RESULTS

Cloning and sequence analysis of *atonal* homologs

The *Drosophila atonal* gene belongs to the family of bHLH transcription factors, but the sequence and size of the bHLH domain are sufficiently different to make it distinctive from other members of this family (Fig. 1). Degenerate PCR primers corresponding to sequences within the basic and the second helix domains of the *Drosophila atonal* protein (13) were used to amplify fragments from genomic DNA of *Tribolium castenium* (red flour beetle), *Fugu rubripes* (puffer fish), and chicken. Sequencing of the cloned PCR products revealed a high degree of similarity between the bHLH domain of the *Drosophila* Atonal and the amplified homologous domain from the various species (67–78% identity). Notably, the cross-species similarity of *atonal* homologs within the bHLH domain was higher than the similarity between *atonal* and other *Drosophila* bHLH-containing genes (20–50%), and all the PCR products had the exact same length as *atonal* (Fig. 1A).

The PCR-generated chicken bHLH fragment was used as a probe to screen a chicken genomic DNA library and several positive clones were identified and partially sequenced. Fragments from these chicken clones were subsequently used as probes to screen a mouse genomic DNA library. Three independent mouse genomic clones were identified and characterized. Sequencing of the bHLH domain in mouse revealed that the highest degree of homology was to the *Drosophila atonal* gene (67% identity) (Fig. 1A). Moreover, the mouse and chicken bHLH domains were almost identical at the protein level (95%).

Mouse-specific oligonucleotides were used to analyze by PCR the expression of *atonal* homolog in four mouse cDNA libraries from embryonic days: E10.5, E12.5, E14.5 and E16.5. Signals were detected in all four libraries. Screening of E10.5 and E12.5 mouse cDNA libraries with the mouse bHLH fragment resulted in the isolation of five independent cDNA clones from the E10.5 library, and only one clone from the E12.5 library. Sequence analysis of the cDNA clones identified the open reading frame (ORF) of the mouse *atonal* homolog, *Math1*. The sequences from genomic and cDNA clones were co-linear, suggesting that *Math1* coding region consists of a single exon. While this work was in progress, the mouse homolog was identified independently by Akazawa *et al.* (16).

To identify the human homolog of *Math1* a human genomic DNA library was screened with a *Math1* probe encompassing the ORF. Sequence analysis of the human homolog, *HATH1*, revealed that the coding region is intronless (Fig. 2). The predicted HATH1 protein has almost the same size as the mouse protein (HATH1 is three amino acids shorter than Math1). HATH1 consists of 354 amino acids, and has a calculated relative molecular weight of 37.8 kDa. Sequence similarity between *HATH1* and *Math1* is very high in the bHLH domain as well as throughout the entire coding region (86 and 89% identity at the nucleotide and amino acid levels, respectively). The high similarity extends to the 5′ and 3′ untranslated sequences (79 and 83% of nucleotide identity, respectively). In contrast, the similarity between the chicken and mammalian homologs is restricted to the bHLH domain. The bHLH domain from chicken is 97 and 95% identical to that of human and mouse, respectively, while the sequences flanking the bHLH domain are only 41 and 46% identical to human and mouse, respectively. It should be noted that the bHLH domain of Math2 (17) is less similar to that of *Drosophila* Atonal (51% identity) compared to the domain of Math1 (67% identity). The bHLH of Atonal and its homologs is identical in length, in contrast to the domain from other *Drosophila* genes. The most variable region is the loop connecting the two helixes, which gives to each bHLH a typical length that can be used in addition to overall sequence similarity to establish homology relationships (Fig. 1A). In vertebrate, the bHLH domain is located in the middle of the coding region, while in *Drosophila* it is located at the carboxy terminal (13). It is concluded that given the divergence between the chicken and the mammalian genes, the bHLH domain may be the only functionally significant domain in the proteins (see discussion).

To address the question of homology relationships between *atonal* and its homologs, the putative gene products of Math1 were compared to the protein sequences available in the databanks. The most similar protein was *Drosophila* Atonal, with a smallest sum probability [P(N)] of $6.3e^{-20}$, and the *C. elegans* LIN-32 with P(N) of $5.7e^{-18}$. Similar results were obtained for *HATH1* with P(N) values of $8.2e^{-20}$ and $2.8e^{-18}$ to Atonal and LIN32, respectively. The next *Drosophila* proteins to be identified by the search were TWIST and ACHETE T5, with the much lower P(N) values of $1.7e^{-5}$ and $6.4e^{-4}$, respectively, compared to Math1, and $2.5e^{-6}$ and $3.9e^{-5}$, respectively, compared to HATH1. When the reciprocal search was performed with *Drosophila* Atonal as the query, the first gene product that was identified was Math1, with P(N) of $9.0e^{-20}$, much higher than any other *Drosophila* protein. The homology relationships are further illustrated in Figure 1B by a dendogram showing the degree of similarity between the bHLH from various proteins. It is clearly seen that the bHLH domains cloned in this study are most similar to Atonal. The high scores obtained for the similarity between Atonal, Math1 and HATH1 provide further support for the claim that all these genes are homologs, although the data does not prove that they are real orthologs.

Expression patterns of *atonal* in chicken and mouse

To analyze the temporal expression of *Math1* transcripts, a developmental mouse embryonic northern analysis was performed. A fragment of *Math1* spanning the bHLH domain was hybridized to poly(A)$^+$ RNA isolated from mouse embryos at different stages of development. As shown in Figure 3, a single 2.5 kb band was detected from E11 to E17 with the strongest signal at E11.

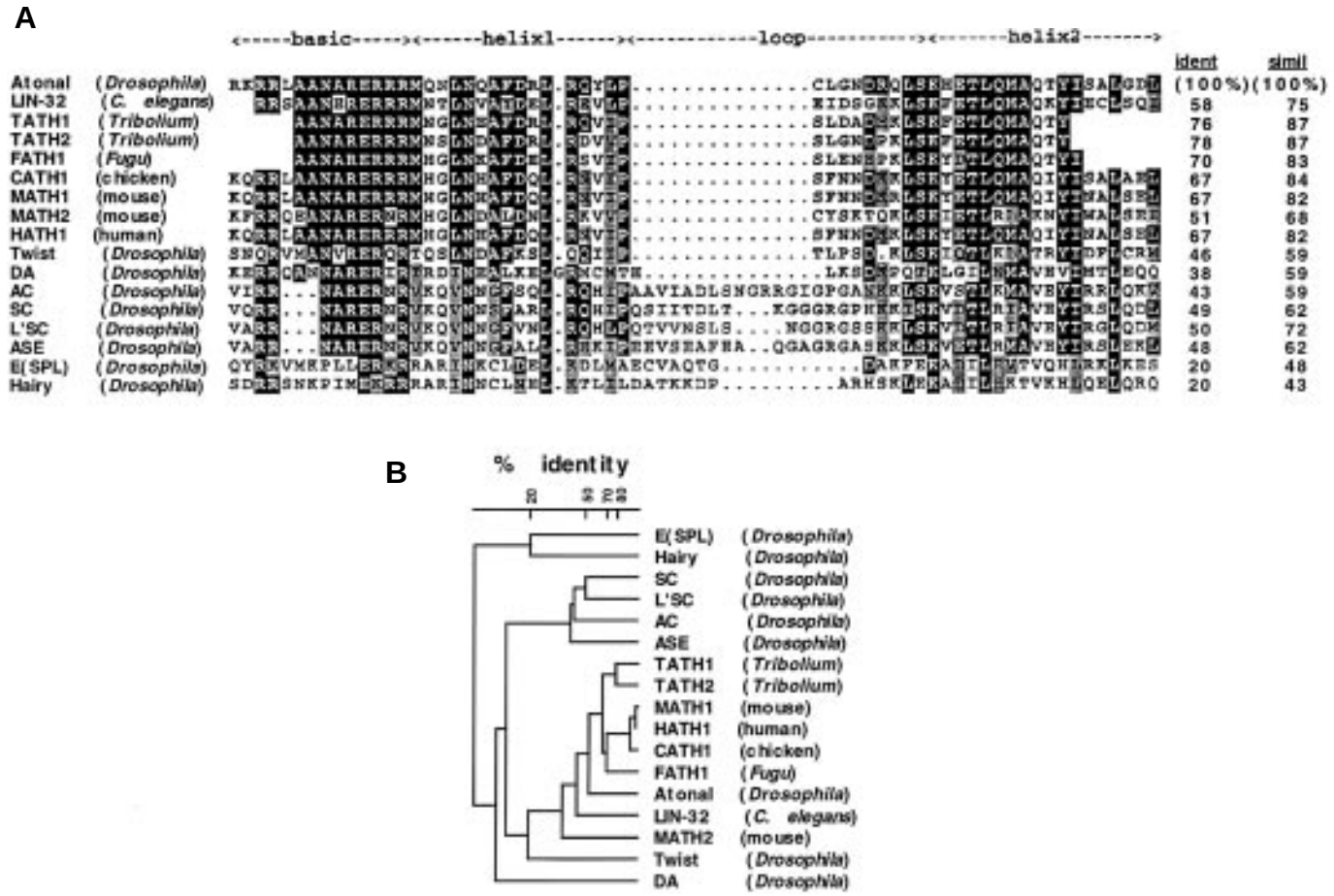


Figure 1. Comparison between Atonal, related sequences from other species and other bHLH-containing *Drosophila* genes. (**A**) Amino acid sequence comparison between the bHLH domains of *Drosophila* Atonal, its homologs and other fruit fly gene products. Shown are the aligned amino acid sequences of the *Drosophila* Atonal bHLH domain (13), the closest homolog cloned so far from *C. elegans*, LIN-32 (38), two isoforms cloned by us from *Tribolium castenium* (red flour beetle) (*Tribolium atonal* homologs, TATH1 and TATH2), *Fugu rubripes* (puffer fish atonal homolog, FATH1), chicken (CATH1), mouse homologs 1 (Math1) and 2 (Math2) (17), and human (HATH1). To emphasize the similarity between Atonal and the putative homologs, included are representative bHLH domains encoded by other *Drosophila* genes [*twist, daughterless, achaete, scute, lethal of scute, asense, enhancer of split,* and *hairy* (adapted from ref. 13)]. These gene products were chosen because they are the most similar to Atonal. Black boxes represent identical residues compared to Atonal and gray boxes denote conservative substitutions. The percentages of identity and similarity to Atonal are given. The Atonal bHLH domain from *Drosophila* is more similar to the domain of the putative homologs from other species, than to the bHLH of other *Drosophila* genes. The bHLH domain from vertebrates has the highest degree of similarity. The human homolog is more similar to *Math1* than to *Math2*, and was therefore designated *HATH1*. (**B**) Identity dendogram of the bHLH domains shown in (A). All protein sequences were aligned and their degree of similarity illustrated by the level of branching. This analysis illustrated further that Atonal is more similar to its putative homologs than to other bHLH-containing genes.

In situ hybridization analysis was used to compare the spatial and temporal expression patterns of mouse and chicken *atonal* homologs (*Math1* and *Cath1*, respectively) in the CNS. In the mouse, expression is first seen by embryonic day 9 in the dorsal part of the neural tube in neuroblasts situated lateral to the roof plate (Fig. 4A, panel 1). Serial sections along the anteroposterior axis of an E12 mouse embryo revealed expression of *Math1* in the dorsal neuroblasts, both in the metencephalon (Fig. 4A, panel 2) and along the entire length of the spinal cord (Fig. 4A, panels 3 to 10), with the exception of the posterior-most region of the spinal cord (data not shown). There is strong expression of *Math1* in the rhombic lip (Fig. 4A, panel 3), the region in which the precursors of the external granule neurons of the cerebellum arise. External granular cells of the postnatal cerebellum strongly express *atonal* in the mouse (16). Similarly, *Cath1* expression in the chick brain was detected between E12 and E18, when maximal granule cell migration occurs (not shown). *Cath1* expression in the chicken spinal cord is identical to that seen in mouse (Fig. 4B, panels 1–6). Expression is first seen around Hamburger-Hamilton stage 17 (E3) in the rostral section of the neural tube and persists until stage 28 (E6, arrows in Fig. 4B, panel 6). Panels 1–3 depict expression in the spinal cord of a stage 22 embryo at the cervical (Fig. 4B, panel 1), lumbar (Fig. 4B, panel 2) and caudal (Fig. 4B, panel 3) levels. Similar to *Math1* expression pattern, *Cath1*-positive cells extend to the roof plate, but do not include this structure (Fig. 4B, panels 4 and 6).

The nature of the *atonal* homologs-positive cells in the spinal cord is presently unknown. However, several molecular markers delineating cell populations in the dorsal region of the neural tube have recently been identified (18,19). Figure 4C depicts the expression of *Math1* (panels 1 and 4), the LIM/homeodomain gene (20), *LH-2*, (panels 2 and 5) and *wnt-1* (21) (panels 3 and 6) mRNA in the caudal (top panels) and cervical regions (bottom panels) of a day 12 mouse embryonic spinal cord. *LH-2* defines a subset of dorsal commissural neurons (18) and *wnt-1* expression defines the dorsal midline of the spinal cord (21). Initially, *Math1* transcripts are found in the dorsal-most region that is adjacent and partially included in the *wnt-1* expression domain (compare

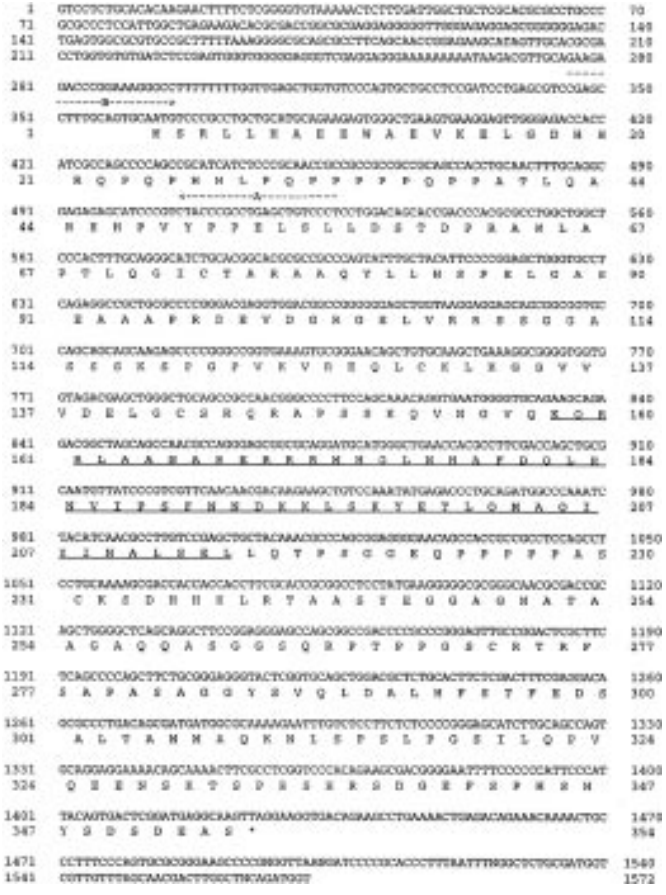


Figure 2. Sequence of *atonal* human homolog, *HATH1*. The nucleotide and deduced amino acid sequences of the human *atonal* homolog gene are presented. In frame stop codon is marked by an *. The sequence of STS UT6525 found under GenBank accession number L30585, is 98% identical to *HATH1* (nucleotides 245–901), and the primers developed for this STS are shown as arrows. The basic region, the two helices and the connecting loop are designated.

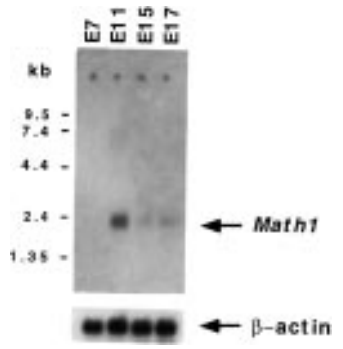
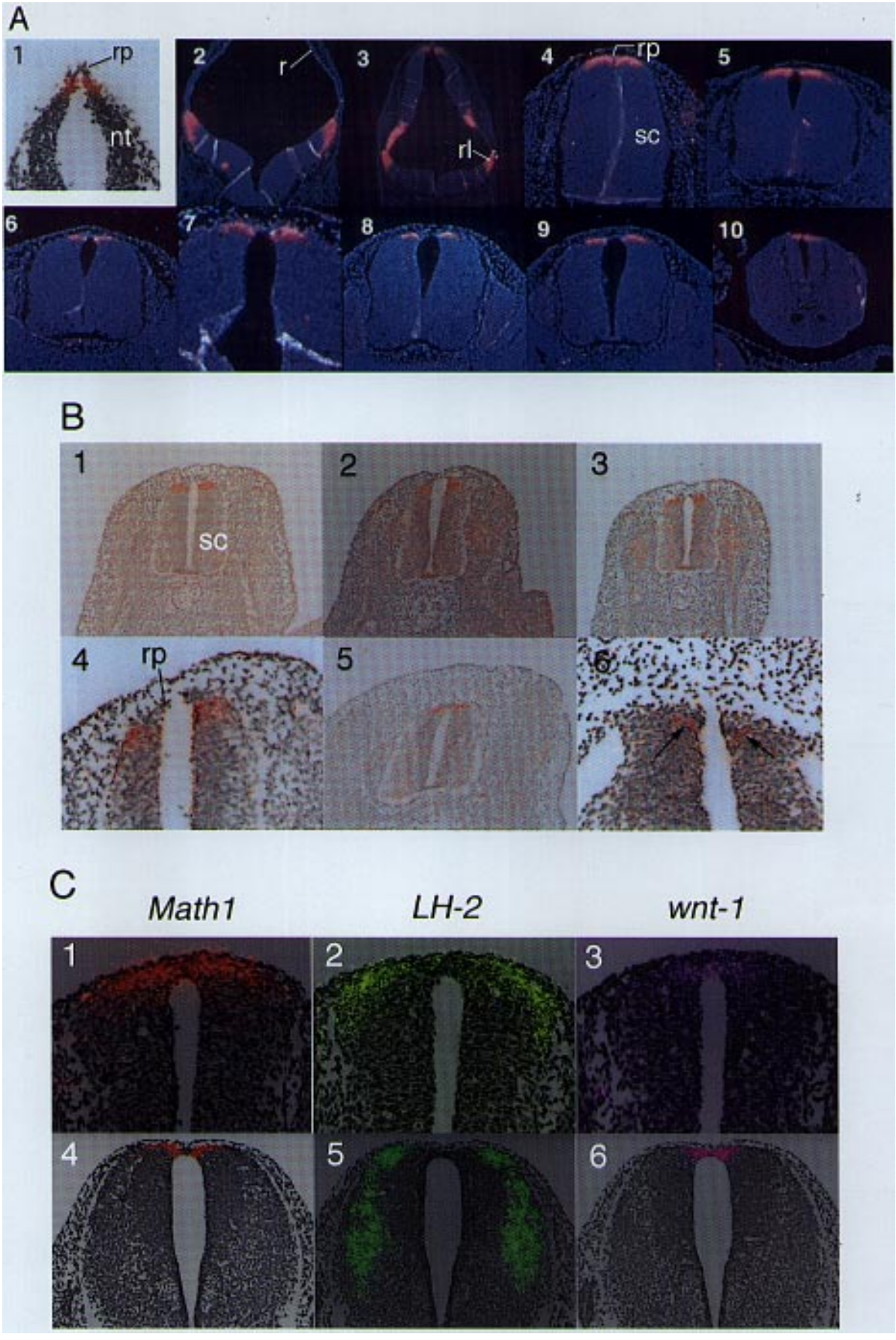


Figure 3. Expression of *atonal* during mouse embryonic development. A northern blot was probed with a fragment of *Math1* spanning the bHLH domain. In each lane the blot contains 2 μg of poly(A)$^+$ RNA, extracted from whole mouse embryos at different stages of embryogenesis (Stratagene). A single *Math1* transcript of 2.5 kb is detected at E11, E15 and E17. Peak expression is noted at E11. Hybridization with a β-actin probe is shown as a control for loading variation.

panels 1 and 3 in Fig. 4C). Likewise, there is an overlap of the expression domains of *Math1* and *LH-2* (Fig. 4C, panels 1 and 2). Since RNA *in situ* hybridization does not afford cellular resolution, it is not possible to be certain whether any of the hybridizing cells simultaneously express *Math1* and *LH-2*. However, the expression patterns establish that cells expressing these two genes are at least intermingled in the more medial domain of the dorsal spinal cord. It is interesting to note that there are *LH-2* positive cells in the lateral region of the spinal cord which does not express *Math1*. This may imply that *LH-2* expression does not depend on the expression of *Math1* after migration has taken place. In the developmentally more advanced cervical region of the spinal cord, *wnt-1* and *Math1* expression still partially overlap (Fig. 4C, panels 4 and 6). In contrast, *Math1* expressing cells and *LH-2* expressing cells localize to different but adjacent regions, with *LH-2* positive cells located lateral to those containing *Math1* mRNA. Taken together, these expression studies indicate that *Math1* is transiently expressed in both chick and mouse in a population of neuroblasts residing lateral to the roof plate. Unlike *LH-2*-expression, which is seen in neuroblasts that migrate laterally (Fig. 4C, panels 2 and 5), *Math1*-expression is seen only in neuroblast that are immediately below the roof plate and are situated more medially. The partial overlap of the expression regions of *Math1* and *LH-2* in the early stages of spinal cord development (Fig. 4C, panels 1 and 2) raises the possibility that *Math1*-expressing neurons define a stem cell population which serves as a precursor of *LH-2* neurons and possibly of other neurons. The evolutionarily conserved localization of *Math1* and *Cath1*-positive cells and the transient expression of these genes, are consistent with this hypothesis.

Figure 4. Expression of *atonal* in chicken and mouse. (**A**) Expression pattern of *Math1* in the spinal cord during early mouse development. (Panel 1) In the spinal cord of E9 embryo, *Math1*-expressing neuroblasts are found lateral to the roof plate which defines the dorsal midline. (Panels 2–10) E12 embryo sectioned perpendicular to the anteroposterior axis. Panel 2 shows the anterior-most section, whereas panel 10 shows a section in the caudal region. (Panels 2, 3) Expression of *Math1* in neuroepithelium immediately lateral to the roof of the fourth ventricle. The section shown in panel 3 passes through the spinal cord (top) and through the metencephalon at the level of the rhombic lip. Note *Math1* expression in the rhombic lip region. Panels 4–10 show the typical dorsal patch of *Math1*-expression in the spinal cord in a region lateral to the roof plate. Abbreviations: nt, neural tuber; r, roof of fourth ventricle; rl, rhombic lip; rp, roof plate; sc, spinal cord. (**B**) Expression pattern of *Cath1* in the spinal cord of the chick. Panels 1–4 represent transverse sections through the dorsal region of a Hamburger-Hamilton stage 22 chick embryo. Panel 1: transverse section at the wing level. Panels 2, 4: transverse section at the lumbar level; panel 4 is a two-fold magnification of panel 2 and illustrates that *Cath1* is expressed in neuroectoderm lateral to the roof plate. Panel 3: transverse section at the caudal level. Panels 4 and 5 represent transverse sections at the wing level of stage 24 and 28 embryos, respectively. Note the marked reduction of *Cath1* by stage 28 (panel 6). (**C**) Expression in the spinal cord of *Math1* (panels 1, 4), *LH-2* (panels 2, 5), and *wnt-1* (panels 3, 6) at the caudal (top row) and cervical (bottom row) levels. Adjacent sections of a E12.5 mouse embryo are shown. Since neurogenesis proceeds in a rostrocaudal fashion, the caudal sections represent a developmentally earlier stage than sections at the rostral level. *Wnt-1* defines the dorsal midline. In the more advanced anterior spinal cord, *LH-2*-expressing neurons have migrated laterally. Note a partial overlap of the expression pattern of the three genes in the caudal region. A partial overlap of the *Math-1* and *wnt-1* expression domains is also seen at the cervical level. Cells expressing *Math-1* remain in a dorsal position throughout spinal cord development.


A
1
rp
nt
2
r
3
rl
4
rp
sc
5
6
7
8
9
10
B
1
sc
2
3
4
rp
5
6
C
Math1
LH-2
wnt-1
1
2
3
4
5
6

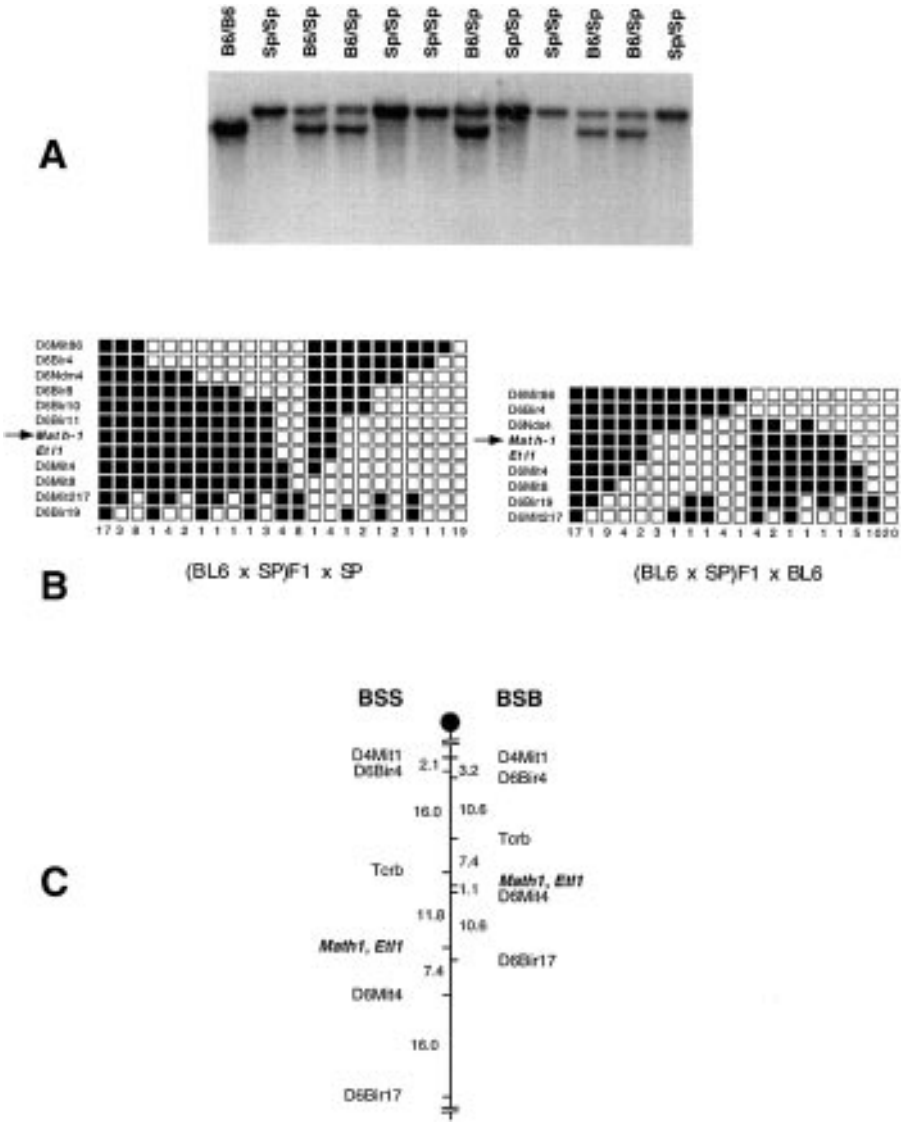


Figure 5. Genetic mapping of *Math1* to mouse chromosome 4. (**A**) Polymorphism in the *Math1* locus between C57BL/6J and SPRETUS/Ei mouse strains. Genomic DNA from interspecific backcross animals [(C57BL/6J × SPRETUS/Ei)F1 × SPRETUS/Ei and (C57BL/6J × SPRETUS/Ei)F1 × C57BL/6J] was digested with *Bam*HI and hybridized to a *Math1* probe. Restriction fragment length polymorphism between the two was identified and used to genotype the animals. (**B**) Distribution of *Math1* and flanking markers in progeny from interspecific backcross mice. Block diagrams of the meiotic breakpoints critical for *Math1* mapping obtained from interspecific backcross of C57BL/6J × SPRETUS/Ei (BSS, left) and C57BL/6J × C57BL/6J (BSB, right). Closed boxes represent the C57BL/6J allele and open boxes the SPRETUS/Ei allele. The numbers of the progeny carrying each type of chromosome are listed at the bottom. (**C**) A schematic genetic map of mouse chromosome 6. Data obtained from the interspecific backcross panels and the cumulative locus map was used to generate the genetic map. *Math1* was mapped between *D6Bir11* and *D6Mit4* markers, without recombinations with the *Etl1* marker, to the same interval where *Lc* and *chp* map. Shown are the localization of *Math1* relative to other markers and the genetic distances between them in cM, based on the BSB and BSS panels.

Mapping of the mouse and human *atonal* homologs

The high conservation of Atonal bHLH domain in many species, in addition to the similar expression patterns in the nervous system of chicken and mouse, suggest that *atonal* homologs have an important function in neuronal development in most species. To determine whether a mouse with *Math1* mutation(s) exists, we mapped this gene in the mouse. The genetic mapping was performed using interspecific backcrosses between C57BL/6J and *Mus spretus* (SPRET/Ei). Hybridization of the *Math1* probe to genomic DNA from the two strains identified a *Bam*HI restriction fragment length polymorphism (Fig. 5A). This polymorphism was subsequently used to genotype 93 animals from a (C57BL/6J × SPRET/Ei)F1 × C57BL/6J backcross, and 92 animals from a (C57BL/6J × SPRET/Ei)F1 × SPRET/Ei backcross.

Analysis of the *Math1* genotype in progeny from the interspecific backcross mice identified recombination events that allowed the regional localization of this gene (Fig. 5B). *Math1* maps to the mid portion of mouse chromosome 6 and is flanked by markers *D6Bir11* and *D6Mit4* (Fig. 5B,C). The size of this interval was determined previously to be approximately 3.5 cM (22). To refine our mapping we genotyped the same backcrosses with a probe derived from the mouse enhancer trap locus 1 (*Etl1*),

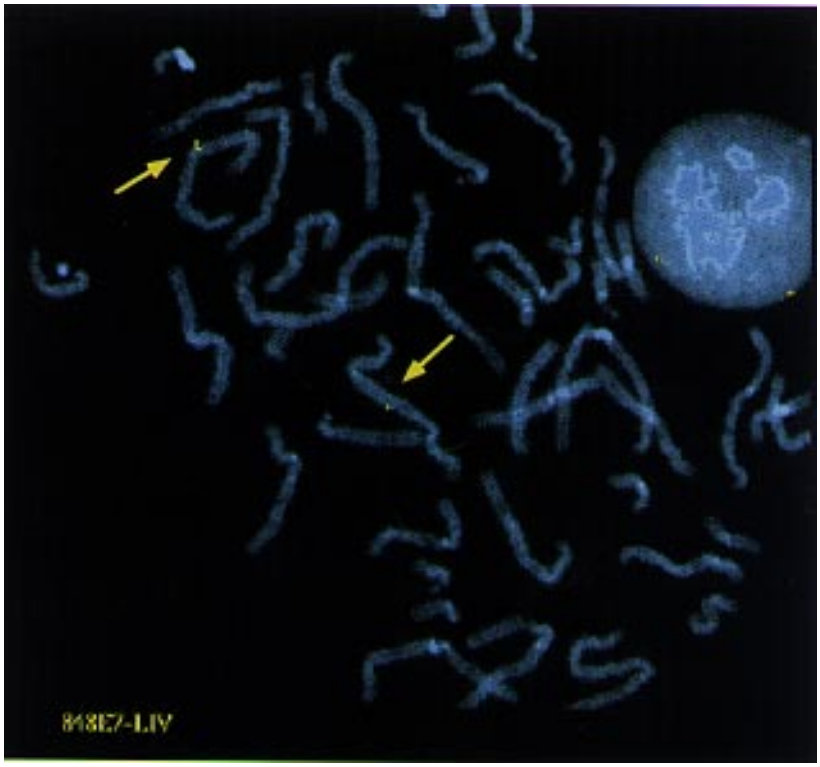


Figure 6. Mapping of *HATH1* to human chromosome 4q22 by FISH. *Alu*-PCR products from a YAC containing *HATH1* were used as a probe on human chromosomes. A consistent signal was obtained only from chromosome 4q22.

which was mapped previously to the mid portion of mouse chromosome 6 between *D6Bir8* and *D6Mit4* (23). *Etl1* encodes a nuclear protein that belongs to the SNF2 superfamily of proteins. It shows a widespread expression pattern throughout development, with particularly high levels present in the Purkinje cells of the cerebellum and might be involved in gene regulation and/or influence chromatin structure (23,24). No recombinations between *Math1* and *Etl1* were found in 192 animals, indicating that the two loci are closely linked. Two neurological mutants, lurcher (*Lc*) (25) and cerebellar-hypoplasia (*chp*) (Dr Muriel Davisson, pers. comm.) have been mapped to the same region. Therefore, the sequence of *Math1* was analyzed in the two mutants, as described below.

To map the human homolog of *atonal*, *HATH1*, we utilized a monochromosomal human-rodent somatic cell hybrid panel. By hybridization of *Hin*dIII-digested DNA from this panel with a probe spanning the bHLH domain, *HATH1* was detected in the hybrid containing human chromosome 4. In addition, we have identified in the GenBank database a human STS, UT6525, that is identical to *HATH1*. The primers and conditions developed to amplify this STS were used to screen the CEPH Mega-YAC library by PCR. Five YACs (673E5, 848E7, 705E11, 877A10, 627A12) were isolated and were shown by Southern analysis to contain the human locus. Searching the genomic databases with these YACs revealed that they map to human chromosome 4, and that one of them contains the dinucleotide repeat marker *D4S1557* (26). This marker maps between *D4S1558* and *D4S411* which are cytogenetically localized to 4q13.3–q21.3 and 4q24–25, respectively (27). *Alu*-PCR products of two YACs were used as probes for fluorescent *in situ* hybridization (FISH) on human metaphase spreads and hybridized to chromosome 4q22 (Fig. 6). To date, no neurological disease is known to map to this cytological band. However, the mapping of *HATH1* within less than one megabase of *D4S1557* makes this marker ideal for genotypic analysis of candidate disorders.

Search for *Math1* mutations in *chp* and *Lc* mutant mice

The *Drosophila atonal* gene product plays an important role in the neurogenesis of the chordotonal organs that are involved in proprioception. In addition, the *atonal* homologs were found to be exclusively expressed in subsets of neurons in the cerebellum, hind brain and spinal cord in both chicken and mouse. The finding that *Math1* maps close to two neurological mutations, *Lc* (25) and *chp* (Dr Muriel Davisson, pers. comm.) both of which cause ataxia, raised the possibility that *Math1* may be disrupted in either mutant mouse. Therefore, the coding region of *Math1* was PCR-amplified, cloned, and sequenced from affected *Lc* and *chp* homozygotes and the appropriate wild type control animals. No mutations were found in the *Math1* gene of the mutant mice throughout the coding region. To identify potential genomic rearrangement in the *Math1* locus, Southern analysis was performed, using various fragments from *Math1* coding and non-coding regions and genomic DNA from affected homozygotes, heterozygotes and wild type mice. No genomic rearrangements were detected but the possibility still remained that mutations in regulatory regions, such as the promoter, may cause the phenotype. Therefore, a genetic analysis was performed to determine more accurately the genetic distance between *Math1* and *chp* or *Lc*. The *Math1* ORF was used as a probe on panels of

interspecific backcrosses. Recombinations were found between *chp* and *Math1* (Drs Susan Cook and Muriel Davisson, pers. comm.), as well as between *Lc* and *Math1* (Dr Nathaniel Heintz, pers. comm.), which confirmed the conclusion that *Math1* is not the gene mutated in these mice.

DISCUSSION

In this study we have shown that the bHLH domain encoded by the *Drosophila atonal* homologs is conserved throughout the animal kingdom. This domain was identified in insects, fish, birds, and mammals and shares high sequence similarity. Homology searches consistently identified the vertebrate sequences as the closest to the fruit fly gene, with much higher scores obtained between Atonal and the vertebrate homologs than between Atonal and any other *Drosophila* gene products. In addition, the size of the bHLH domain is the same in *atonal* homologs from all species tested. Since other bHLH-containing *Drosophila* genes have varying size, due to differences in the loop length, the unique size found in Atonal homologs may be used as another criterion to establish the homology relationship. The conservation of the sequence and the size of the bHLH domain from *Drosophila* Atonal and the putative homologs from other species suggests that the cloned genes are the homologs of *atonal*. The bHLH domain is almost identical in vertebrates, while the whole coding region and its flanking sequences are highly similar in mammals. It was shown (13) that the bHLH domain is necessary and sufficient to promote correct protein dimerization and recognition of the DNA target, and might therefore be considered the 'core' of the protein. The high homology between the *Drosophila* and vertebrate Atonal bHLH domain suggests that the function of the protein is conserved in different species.

Similar conservation of the bHLH domain was found also for the genes of the *achaete-scute* complex (AS-C). The similarity in the AS-C genes between different species is very akin to our findings. In both cases, animals that are evolutionarily distant show sequence and size conservation only at the bHLH domain whereas mammalian homologs are conserved throughout the coding region. Functional conservation was shown for the hydra homolog of AS-C through its proneural activity in *Drosophila* (28). This may be due to the fact that although the overall identity between the hydra and the *Drosophila* homologs is low, the bHLH domain is 67% identical. The same conservation of function may apply to the *atonal* homologs, which have low overall similarity with *Drosophila atonal*, but high degree of similarity at the bHLH domain (e.g. in chicken and mouse). The demonstration of conservation of neuronal expression patterns in mouse and chicken homologs of *atonal* in this study, in addition to their sequence conservation, suggests a preservation of a function during the early stages in the development of the nervous system.

The vertebrate *atonal* homologs were shown by RNA *in situ* hybridization to be expressed in mouse starting at E9 and in chicken by Hamburger-Hamilton stage 17. Both homologs are expressed in the dorsal-most part of the spinal cord and hind brain, lateral to the roof plate along the entire length of the spinal cord. There is also strong expression of *Math1* and *Cath1* in the rhombic lip, which contain the cells that will give rise to the external granular layer of the cerebellum. The rhombic lip expresses *Math1* and *Cath1* at the stage in which the granular cell precursors are being born, prior to their dorsomedial migration, a process that eventually produces the external granular layer (29). Our expression studies raise the possibility that the dorsal neuroectodermal cells which express *atonal* give rise to the *LH-2* positive commissural neurons. However, the possibility remains, that neuroepithelial cells in the spinal cord will undergo a second migration wave, through the developing dorsal roots, after the emigration of neural crest cells has been completed, to acquire various fates (30). Thus in vertebrates, *atonal* homologs might be involved in the specification of the fate of neuronal precursor cells, as was shown in *Drosophila*. However, in *Drosophila* the progenitors are ectodermal cells, while in vertebrates the progenitors may be already committed to a neuronal fate or may even be neurons already at the time *atonal* homologs are expressed.

It seems feasible to hypothesize that mutations in *Math1* may affect the development of the mouse nervous system based on the role of *atonal* in the development of the *Drosophila* chordotonal organs, the fact that adult flies that lack *atonal* have poor coordination (15), and the expression patterns in the brain and the spinal cord of vertebrates. A search for mouse and/or human disease caused by mutation in *Math1* and *HATH1*, respectively, was undertaken through their mapping. The mapping of *atonal* homologs to mouse chromosome 6 and to human chromosome 4q22 defines a new linkage conservation between the two species. Genes more centromeric to *Math1* map to human chromosome 7, while the region that is more distal on mouse chromosome 6 maps to human chromosome 2.

It was intriguing to discover that two neurological mutants *chp* and *Lc* map within 2 cM of *Math1*. We did not find any mutations in the coding region of *Math1* or any genomic rearrangements in homozygote *Lc/Lc* or *chp/chp* mutant mice. Furthermore, recombinations between *Math1* and both *chp* and *Lc*, excluded *Math1* as a candidate gene for these mutants. A more direct approach to reveal the function of *Math1* in the development of the nervous system is the generation of mice that lack the gene, which is currently in progress (N.B.A. and H.Y.Z., unpublished results).

The identification of the human homolog, *HATH1*, and its mapping to human chromosome 4q22 will allow the testing of patients with unmapped inherited neurological disorders, especially those manifesting ataxia. The fact that the coding region of *HATH1* is encoded in one relatively small exon will allow the direct amplification and sequence analysis of *HATH1* from patients. In addition, the use of the adjacent polymorphic marker, *D4S1557*, may facilitate such a screen.

To conclude, we have presented data that show that a cross species conservation of a *Drosophila* proneural gene exists throughout evolution. The fact that both sequence and expression pattern similarities are identified for *atonal* homologs may imply that similar, even if not identical, pathways for the early development of the nervous system may occur in insects, birds and mammals.

MATERIALS AND METHODS

Cloning of *atonal* homologs

Degenerate primers corresponding to the basic and second helix of *Drosophila atonal* (13) were designed: D2, residues 1109–1128 (GCIGCIAA(C/T)GCI(A/C)GIGA(A/G)(C/A)G) and D1, residues 1226–1246 (IAT(G/A)TAIGT(C/T)TGIGCCAT(C/A)TG). PCR-amplification was carried out in 10 mM Tris, pH 8.3, 50 mM KCl, 0.01% gelatin, 1.25 mM $MgCl_2$, 0.25 mM each dNTP, 1.25 U AmpliTaq (Cetus)/50 µl and 1 µM of each

primer. Amplification was performed in MJR thermocycler, programmed as follows: 40 cycles of denaturation at 94°C for 1 min, annealing at 50°C for 1 min and elongation at 72°C for 1 min. Sequencing was performed either manually, using Sequenase 2.0 kit (USB) or on an ABI 373A sequencer with dyedeoxy terminators. Homology searches were performed on the National Center for Biotechnology Information (NCBI) server by the BLASTP program (31), using BLOSUM62 matrix and no filters. Databases searches included GenBank, PDB, SwissProt and PIR.

The following libraries were screened or utilized for PCR-amplification: *Tribolium castenium* embryonic cDNA library (kindly provided by Dr Sue Brown), Chicken embryonic (HH14–17) cDNA library (Stratagene), Mouse embryonic (E10.5, E12.5, E14.5, E16.5) cDNA libraries (kindly provided by Dr Allen Bradley), Mouse 129/Sv genomic library (Stratagene), and Human genomic library (kindly provided by Dr Christie Ballantyne). The libraries were screened according to standard procedures (32). For cross species hybridization, the filters were washed at low stringency (3× SSC at 50°C; 1× SSC is 0.15 M NaCl and 15 mM citric acid), while for cross mammals the hybridization stringency was higher (0.1–0.5× SSC at 65°C).

Genetic mapping of *Math1*

To map *Math1* genomic DNA from C57BL/6J and SPRETUS/Ei mouse strains was digested with a battery of restriction enzymes and hybridized to a *Math1* probe. *Bam*HI digestion produced the largest visible difference between the two, and was thus chosen for the genotyping of the animals. Two backcross panels, BSS and BSB (The Jackson Laboratory), were used to identify the critical recombinations that define the interval at which *Math1* maps (see Results).

Physical mapping of *HATH1*

To identify human YACs containing *HATH1*, we used the primers and conditions developed for STS UT6525 (GenBank accession number L30585), which was found by homology searches to be identical to *HATH1*. These primers were utilized to screen the CEPH Mega YACs library by PCR. DNA was extracted from the positive YACs in agarose blocks (33), and subjected to *Alu*-PCR using PDJ34 primers (34). FISH was performed on human chromosomes as described in (35). Chromosomes were counterstained with 4′,6-diamidin-2-phenylindol-dihydrochlorid (DAPI) to determine the subchromosomal localization of the signal. More than 10 chromosomes were examined for each probe.

Math1 mutation analysis

To identify mutations in the open reading frame of *Math1*, the following primers were used for PCR amplification: MA5 (ACCTCCTCTAACACGGCAC) and MA6 (AGGGCATTTG-GTTGTCTCAG). PCR was performed in a Cetus 9600 thermo-cycler, programmed as follows: 35 cycles of denaturation at 94°C for 30 s, annealing at 60°C for 30 s and elongation at 72°C for 100 s, with an extension of the first and last steps. Amplification mix was as above, except for a $MgCl_2$ concentration of 3 mM. PCR products were subcloned into pBlueScriptII/KS$^+$ (Stratagene) and sequenced on an ABI sequencer. To achieve high accuracy and overlapping readings, the following internal oligonucleotides were used for sequencing, in addition to MA5 and MA6: MA1 (CAGAAGCAAAGGAGGCTGGC), MA2 (GCTTCTTGTCGTTGTTGAAGG), MA3 (TCTGCTGCATT-CTCCCGAGC) and MA4(GCACCGAGTAACCCCCAGAG).

As templates for mutation analysis we have used genomic DNA extracted from *Lc/Lc* and *chp/chp* mutant mice. *Lc/+* mice are maintained at the Jackson Laboratory by mating with (C57BL/6J-AwJ × CBA/Ca)F1 mice. To obtain *Lc/Lc* homozygotes, *Lc/+* mice were intercrossed. *Lc/+* mice are ataxic, while *Lc/Lc* mice die within a few hours of delivery (36). DNA from *chp/+* and *chp/chp* mice was obtained from Dr Muriel. Davisson. Control DNA was from C3H/HeJ-dw[J]/+ mice, the inbred strain on which *chp* mutation occurred. The same DNA samples were used in Southern analysis (32).

Northern analysis

Mouse developmental northern blot was purchased from Stratagene. In each lane, the blot contains 2 µg of poly(A)$^+$ RNA extracted from whole mouse embryos at different stages of development. Hybridization was performed according to the manufacturer's instructions using a *Math1* PCR product spanning the bHLH domain as a probe.

In situ hybridization

Embryo collection, sectioning and *in situ* hybridization were performed as previously described (37). Fragments of chicken and mouse *atonal* homologs in pBSII-KS$^+$ were linearized and used as templates to transcribe either the sense or antisense [^{35}S]-labeled riboprobes. Photographs are double exposures; *in situ* hybridization signals are colored, while the grey or blue color represents the nuclei stained with Hoechst 33258 dye. Both chicken and mouse probe contain the first 60 bp of the bHLH domain. The chicken probe extends 273 bp upstream to the bHLH domain, while the mouse probe includes about 1 kb 5′ to the bHLH domain.

ACKNOWLEDGEMENTS

We thank the YAC core, headed by Dr Craig Chinault, and the FISH core, headed by Dr Antonio Baldini, both at the Genome Center at Baylor College of Medicine, for the screening of the YAC library and FISH mapping, respectively. We are grateful to Dr Eva Eicher for the generous gift of backcross DNA panels, assistance in analysis of the mouse mapping data and for critical review of the manuscript. We would like to thank Dr Muriel Davisson for providing *chp* DNA and sharing unpublished information, Dr Achim Gossler for providing the *Etl1* probe, Dr Nathaniel Heintz for sharing unpublished mouse mapping data, Dr Igna Van den Veyver for comments on the manuscript, and Catherine Tasnier for catalyzing this fruitful collaboration. H.Y.Z. is an Investigator and H.J.B. is an Associate Investigator from the Howard Hughes Medical Institute. This work was supported by a grant from the National Institutes of Health, National Institute of Neurological Disease and Stroke (NS27699) to H.Y.Z.

REFERENCES

1. Joyner, A.L. and Guillemot, F. (1994) Gene targeting and development of the nervous system. *Curr. Op. Neurobiol.*, **4**, 37–42.
2. 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.

3. Bodmer, R., Barbel, S., Shepherd, S., Jack, J.W., Jan, L.Y. and Jan, Y.N. (1987) Transformation of sensory organs by mutations of the *cut* locus of *D. melanogaster. Cell*, **51**, 293–307.
4. Vaessin, H., Ellsworth, G., Wolff, E., Bier, E., Jan, L.Y. and Jan, Y.N. (1991) *prospero* is expressed in neuronal precursors and encodes a nuclear protein that is involved in the control of axonal outgrowth in *Drosophila. Cell*, **67**, 941–953.
5. Kolodkin, A.L., Matthes, D.J. and Goodman, C.S. (1993) The *semaphorin* genes encode a family of transmembrane and secreted growth cone guidance molecules. *Cell*, **75**, 1389–1399.
6. Andres, V., Nadal-Ginard, B. and Mahdavi, V. (1992) *Clox*, a mammalian homeobox gene related to *Drosophila cut*, encodes DNA-binding regulatory proteins differentially expressed during development. *Development*, **116**, 321–334.
7. Oliver, G., Sosa-Pineda, B., Geisendorf, S., Spana, E.P., Doe, C.Q. and Gruss, P. (1993) *Prox 1*, a *prospero*-related homeobox gene expressed during mouse development. *Mech. Dev.*, **44**, 3–16.
8. Salzberg, A. and Bellen, H.J. (1996) Invertebrate versus vertebrate neurogenesis: variations of the same theme? *Dev. Genet.*, **18**, 1–10.
9. Murre, C. and Baltimore, D. (1992) In McKnight, S.L. and Yamamoto, K.R. (eds), *Transcriptional regulation.* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, pp. 861–879.
10. Murre, C., Bain, G., van Dijk, M.A., Engel, I., Furnari, B.A., Massari, M.E., Matthews, J.R., Quang, M.W., Rivera, R.R. and Stuiver, M.H. (1994) Structure and function of helix-loop-helix proteins. *Biochem. Biophys. Acta.*, **1218**, 129–135.
11. Jan, Y.N. and Jan, L.Y. (1994) Genetic control of cell fate specification in *Drosophila* peripheral nervous system. *Ann. Rev. Genet.*, **28**, 373–393.
12. 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. *Crit. Rev. Neurobiol.*, **9**, 177–188.
13. 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.
14. 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.
15. 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.
16. Akazawa, C., Ishibashi, M., Shimizu, C., Nakanishi, S. and Kageyama, R. (1995) A mammalian 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.
17. Shimizu, C., Akazawa, C., Nakanishi, S. and Kageyama, R. (1995) MATH-2, a mammalian 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.
18. Liem, K.F., Tremml, G., Roelik, H. and Jessell, T.M. (1995) Dorsal differentiation of neural plate cells induced by BMP-mediated signals from epidermal ectoderm. *Cell*, **82**, 969–979.
19. Pfaff, S.L., Mendelsohn, M., Stewart, C.L., Edlund, T. and Jessell, T.M. (1996) Requirement for LIM homeobox gene Isl1 in motor neuron generation reveals a motor neuron-dependent step in interneuron differentiation. *Cell*, **84**, 309–320.
20. Xu, Y., M., B., Fisher, P., Rathbun, G., Oltz, E.M., Yancopoulos, G.D., Jessell, T.M. and Alt, F.W. (1993) *LH-2*: a LIM/homeodomain gene expressed in developing lymphocytes and neural cells. *Proc. Natl Acad. Sci. USA*, **90**, 227–231.
21. Parr, B.A., J., S.M., Vassikeva, G. and McMahon, A.P. (1993) Mouse *Wnt* genes exhibit discrete domains of expression in the early embryonic CNS abd limb buds. *Development*, **119**, 247–261.
22. Mouse Genome Database (MGD) 3.1. (1996) Mouse Genome Informatics, The Jackson Laboratory, Bar Harbor, Maine. (URL: http://www.informatics.jax.org/).
23. Darling, S.M., Guenet, J.L., Simon-Chazottes, D., Shiels, P., Davies, W. and Gossler, A. (1996) The enhancer trap locus (*Etl1*) is closely linked to lurcher on mouse chromosome 6. *Mamm. Genome*, submitted.
24. Soininen, R., Schoor, R., Henseling, U., Tepe, C., Kisters Woike, B., Rossant, J. and Gossler, A. (1992) The mouse Enhancer trap locus 1 (*Etl-1*): a novel mammalian gene related to *Drosophila* and yeast transcription regulator gene. *Mech. Dev.*, **39**.
25. Norman, D.J., Fletcher, C. and Heintz, N. (1991) Genetic mapping of the lurcher locus on mouse chromosome 6 using intersubspecific backcross. *Genomics*, **9**, 147–153.
26. Gyapay, G., Morissette, J., Vignal, A., Dib, C., Fizames, C., Millasseau, P., Marc, S., Bernardi, G., Lathrop, M. and Weissenbach, J. (1994) The 1993–94 Genethon human genetic linkage map. *Nature Genet.*, **7**, 246–339.
27. Buetow, K.H., Ludwigsen, S., Scherpbier-Heddema, T., Quillen, J., Murray, J.C., Sheffield, V.C., Duyk, G.M., Weber, J.L., Weissenbach, J., Gyapay, G. and *et al.* (1994) Human genetic map. Genome maps V. Wall chart. *Science*, **265**, 2055–2070.
28. Grens, A., Mason, E., Marsh, J.L. and Bode, H.R. (1995) Evolutionary conservation of a cell fate specification gene: the hydra *achaete-scute* homolog has proneural activity in *Drosophila. Dev.*, **121**, 4027–4035.
29. Hatten, M.E. and Heintz, N. (1995) Mechanisms of neural patterning and specification in the developing cerebellum. *Ann. Rev. Neurosci.*, **18**, 385–408.
30. Sharma, K., Korade, Z. and Frank, E. (1995) Late-migrating neuroepithelial cells from the spinal cord differentiate into sensory ganglion cells and melanocytes. *Neuron*, **14**, 143–152.
31. Altschul, S.F., Gish, W., Miller, W., Myers, E.W. and Lipman, D. (1990) Basic local alignment tool. *J. Mol. Biol.*, **215**, 403–410.
32. Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989) *Molecular Cloning. A Laboratory Manual.* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.
33. Den Dunnen, J.T. and Van Ommen, G.J.B. (1991) In Mathew, C. (ed.), *Methods in Molecular Biology.* The Humana Press Inc., Clifton, Vol. 9, pp. 169–182.
34. Zoghbi, H.Y. and Chinault, C. (1994) In Nelson, D.L. and Brownstein, B.H. (eds), *YAC libraries, A User's Guide.* W.H. Freeman and Co., New York, pp. 93–112.
35. Baldini, A. and Lindsay, E. (1994) In Choo, K.H.A. (ed.), *Methods in Molecular Biology.* Humana Press Inc., Totowa, Vol. 33, pp. 75–84.
36. Heintz, N., Norman, D., Gao, W.-Q. and Hatten, M. (1993) In Davies, K.E. and Tilghman, S.M. (eds), *Genome maps and neurological disorders.* Cold Spring Harbor Laboratory Press, New York, Vol. 6, pp. 19–45.
37. Sundin, O., Busse, H.G., Rogers, M.B., Gudas, L.J. and Eichele, G. (1990) Region-specific expression in early chick and mouse embryos of Ghox-lab and Hox 1.6, vertebrate homeobox-containing genes related to *Drosophila labial. Development*, **108**, 47–58.
38. Zhao, C. and Emmons, S.W. (1995) A transcription factor controlling development of peripheral sense organs in *C. elegans. Nature*, **373**, 74–78.