DEVELOPMENTAL BIOLOGY **197,** 106–116 (1998)
ARTICLE NO. DB988873

Multiple Elements Regulate *Mash1* Expression in the Developing CNS

Sunita Verma-Kurvari, Trisha Savage, Derek Smith, and Jane E. Johnson[1]

Cell Biology and Neuroscience, University of Texas Southwestern Medical Center, 5323 Harry Hines Boulevard, Dallas, Texas 75235-9111

Mash1, a transcription factor of the basic helix–loop–helix class, is expressed during embryogenesis in restricted regions of the nervous system. An essential role for Mash1 in neural development was demonstrated previously in mice carrying a targeted disruption of the *Mash1* gene. Regulation of the precise temporal and spatial expression of *Mash1* is thus likely to be important for proper neural development. In this study, sequences that regulate *Mash1* expression in the central nervous system were characterized by assaying the expression of *lacZ* reporter genes in transgenic embryos. A 1158-bp enhancer localized ~7 kb upstream of the *Mash1* coding region was identified. Deletions within this enhancer region reveal the presence of both positive and negative *cis*-acting elements. Analysis of multiple sequences within the enhancer demonstrate that different elements preferentially function in different regions within the *Mash1*-specific CNS expression domain. In addition, a role for sequences 3′ of the *Mash1* coding region is revealed, providing evidence for posttranscriptional control of *Mash1* expression in multiple CNS domains.

Key Words: **bHLH; neural development; transcription.**

INTRODUCTION

Transcription factors of the basic helix–loop–helix (bHLH) family are involved in the development of multiple cell types including muscle and neuronal lineages in both invertebrates and vertebrates (Kageyama *et al.*, 1995; Yun and Wold, 1996). Expression of members of one subclass within the bHLH family is restricted to early stages of neural development. This subclass includes the *Drosophila* proneural genes of the *achaete–scute* complex and *atonal* that are required for the development of specific neuronal sublineages (Campuzano and Modolell, 1992; Jarman *et al.*, 1993). Related genes in mammals include *Mash1*, *NeuroD*, *neurogenin1/Math4C*, *Math4A/neurogenin2*, *Neurogenin3*, *Math1*, *Nex1/Math2*, and *Math3* (Johnson *et al.*, 1990; Bartholoma and Nave, 1994; Akazawa *et al.*, 1995; Lee *et al.*, 1995; Shimizu *et al.*, 1995; Gradwohl *et al.*, 1996; Sommer *et al.*, 1996; Lemercier *et al.*, 1997; Takebayashi *et al.*, 1997). Each member of this subclass has a distinct pattern of expression relative to both the timing of neuronal differentiation and the domains of expression within the nervous system. *Mash1*, *Math1*, *neurogenin1/Math4C*, and *Math4A/neurogenin2* are expressed in proliferating neuronal progenitors (Lo *et al.*, 1991; Guillemot and Joyner, 1993; Akazawa *et al.*, 1995; Ben-Arie *et al.*, 1996; Gradwohl *et al.*, 1996; Helms and Johnson, 1998). In some regions, *Mash1* and *Math1* appear to be localized to early differentiating cells since their expression overlaps with both mitotic and postmitotic cells (Porteus *et al.*, 1994; Gordon *et al.*, 1995; Helms and Johnson, 1998). In contrast, other members of the neural subclass of bHLH factors are expressed at later stages of neuronal differentiation such as *NeuroD* and *NSCL1/2* in postmitotic neuronal precursors, and *Nex1/Math2* in terminally differentiated cells (Begley *et al.*, 1992; Bartholoma and Nave, 1994; Lee *et al.*, 1995; Shimizu *et al.*, 1995; Kawakami *et al.*, 1996; Kume *et al.*, 1996; McCormick *et al.*, 1996; Yasunami *et al.*, 1996). The role of these transcription factors in neural development is beginning to be elucidated in overexpression and loss of function studies. These studies suggest that the precise temporal and spatial expression of these bHLH factors is critical for proper development of the nervous system (Guillemot *et al.*, 1993; Ferreiro *et al.*, 1994; Turner and Weintraub, 1994; Ishibashi *et al.*, 1995; Lee *et al.*, 1995; Ma *et al.*, 1996; Kim *et al.*, 1997).

Mash1 is the best-characterized member of the vertebrate neural-specific subclass of bHLH factors. It is expressed

[1] To whom correspondence should be addressed. Fax: 214-648-1801. E-mail: Jane.Johnson@email.swmed.edu.

0012-1606/98 $25.00

transiently in the developing nervous system in spatially and temporally restricted domains in the central (CNS) and peripheral (PNS) nervous systems, olfactory epithelium, and the retina (Lo *et al.*, 1991; Guillemot and Joyner, 1993; Porteus *et al.*, 1994; Ahmad, 1995; Gordon *et al.*, 1995; Jasoni and Reh, 1996). In the CNS, Mash1 is localized to domains within the ventral telencephalon, the diencephalon, the dorsal mesencephalon, metencephalon, myelencephalon, and dorsal spinal neural tube. Much lower levels of *Mash1* mRNA expression have also been observed in dorsal telencephalon and ventral mesencephalon (Guillemot and Joyner, 1993). In the PNS, expression is observed in sympathetic, parasympathetic, and enteric neural precursors but not in neural precursors of the sensory, cranial or dorsal root ganglia. *In vitro* studies using neural crest and olfactory epithelial cell cultures suggest that *Mash1* functions in early differentiating neuronal precursors within these lineages (Gordon *et al.*, 1995; Lo and Anderson, 1995; Sommer *et al.*, 1995). An essential role for Mash1 in neurogenesis was demonstrated by a targeted disruption of the *Mash1* gene which results in the loss of most peripheral autonomic and olfactory epithelial neurons (Guillemot *et al.*, 1993).

Given the essential role of Mash1, and the restricted spatial and temporal expression pattern during neurogenesis, it is likely that the precise regulation of *Mash1* expression is critical for proper neural development. The identification of *cis*-acting elements required for *Mash1* expression provides a powerful approach to identify signaling pathways that control early stages of nervous system development. In a previous report, we demonstrated that *cis*-regulatory elements responsible for *Mash1* expression in the CNS reside within 10.5-kb 5′ and 2.7-kb 3′ flanking the coding sequence and that the elements required for expression in the PNS, olfactory epithelium, and retina are likely in distinct regions (Verma-Kurvari *et al.*, 1996). Here, we delineate an enhancer within the *Mash1* 5′ flanking region. This enhancer contains both positive and negative *cis*-acting elements which combine to drive *Mash1*-specific expression in multiple domains within the developing CNS. The data suggest that this enhancer may also contain elements for expression in the PNS and the olfactory epithelium.

MATERIALS AND METHODS

Transgene Construction

Genomic *Mash1* fragments tested in this study were isolated from a mouse pWE15 cosmid library from Stratagene. The assembly of transgenes 1 and 3 have been described previously (Verma-Kurvari *et al.*, 1996, see J1B and J1D). They contain 10.5-kb (No. 1) and 2.8-kb (No. 3) 5′ and 2.7-kb 3′ of the *Mash1* coding region. The *lacZ* reporter construct driven by the *Mash1* basal promoter, transgene 2, is a SphI fragment of 3. It contains 0.9-kb 5′ sequence including 315 bp upstream of the transcription start site and 575 bp of the untranslated region, and 2.7 kb on the 3′ end including the untranslated region, an intron, and the polyadenylation signal (Fig. 1). For assembly of transgenes 4–10, the restriction enzyme fragments indicated in Fig. 1 were cloned 5′ to transgene 2.

Heterologous reporter construct *hsp68lacZpA* (Kothary *et al.*, 1989) was kindly provided by Dr. J. Rossant. All constructs in Fig. 3 place fragments of the 5′ distal *Mash1* region upstream of the *hsp68* basal promoter in *hsp68lacZpA*. For assembly of transgenes 11 and 14, the specified restriction enzymes were used. Transgene 12 was made from No. 5 by replacing a *Kpn*I/*Cla*I fragment with a *Kpn*I/*Cla*I fragment from *hsp68lacZpA* which replaces the *Mash1* proximal 0.9-kb sequence with the *hsp68* basal promoter but retains the 3′ UTR from *Mash1*. Transgene 13 is similar to No. 11 with the addition of the *Mash1* 3′ sequence in the opposite orientation. *Mash1* fragments for transgenes 15–33 were generated by PCR with appropriate primers from within the 1158-bp enhancer sequence. Transgenes generated by PCR were sequenced across the PCR regions.

Further details on generation of constructs will be provided on request. Transgene fragments were isolated from vector sequences and prepared for injection as previously described (Verma-Kurvari *et al.*, 1996).

Generation and Analysis of Transgenic Mice

Transgenic embryos were generated and screened as described previously using eggs from B6D2F1 crosses (Verma-Kurvari *et al.*, 1996). Data for transgene 1 are from established transgenic mice lines. All other transgenes were analyzed as founder embryos. Embryos were harvested at embryonic day E11.5 and fixed in 4% formaldehyde for 30 min at room temperature for β-gal staining as described (Verma-Kurvari *et al.*, 1996). Embryos positive for β-gal activity were postfixed, cleared in methyl salicylate, and photographed under bright field. To identify expression in specific regions, embryos were rehydrated, embedded in agarose, and vibratome sectioned at 300 μm. Transgenic embryos were identified by dot blot analysis of the yolk sac DNA using a ^{32}P-labeled *lacZ* probe (Verma-Kurvari *et al.*, 1996).

RESULTS

Localization of a CNS-Specific Enhancer

Sequences driving *Mash1* expression in specific CNS regions reside within 10.5 kb upstream and 2.7 kb downstream of the *Mash1* coding region (Fig. 2A and Verma-Kurvari *et al.*, 1996). In this study, smaller fragments from the 5′ flanking sequence were tested in transgenic mice to further characterize the *cis*-acting regulatory elements controlling *Mash1* expression in the CNS. Each transgene was injected into fertilized mouse eggs generating transgenic embryos. *LacZ* expression was analyzed in embryos at E11.5, a stage at which high levels of *Mash1* expression are detected in many regions of the CNS and the PNS (Lo *et al.*, 1991; Guillemot and Joyner, 1993). The 5′ fragments were tested for enhancer activity in combination with a *lacZ* reporter construct driven by the *Mash1* basal promoter and containing *Mash1* 3′ sequence (Fig. 1, transgene 2). The basic *Mash1/lacZ* reporter by itself is not sufficient to drive *lacZ* expression in the CNS (Fig. 1, transgene 2). A single 3.3-kb *Sac*I fragment, located more than 5 kb 5′ of the

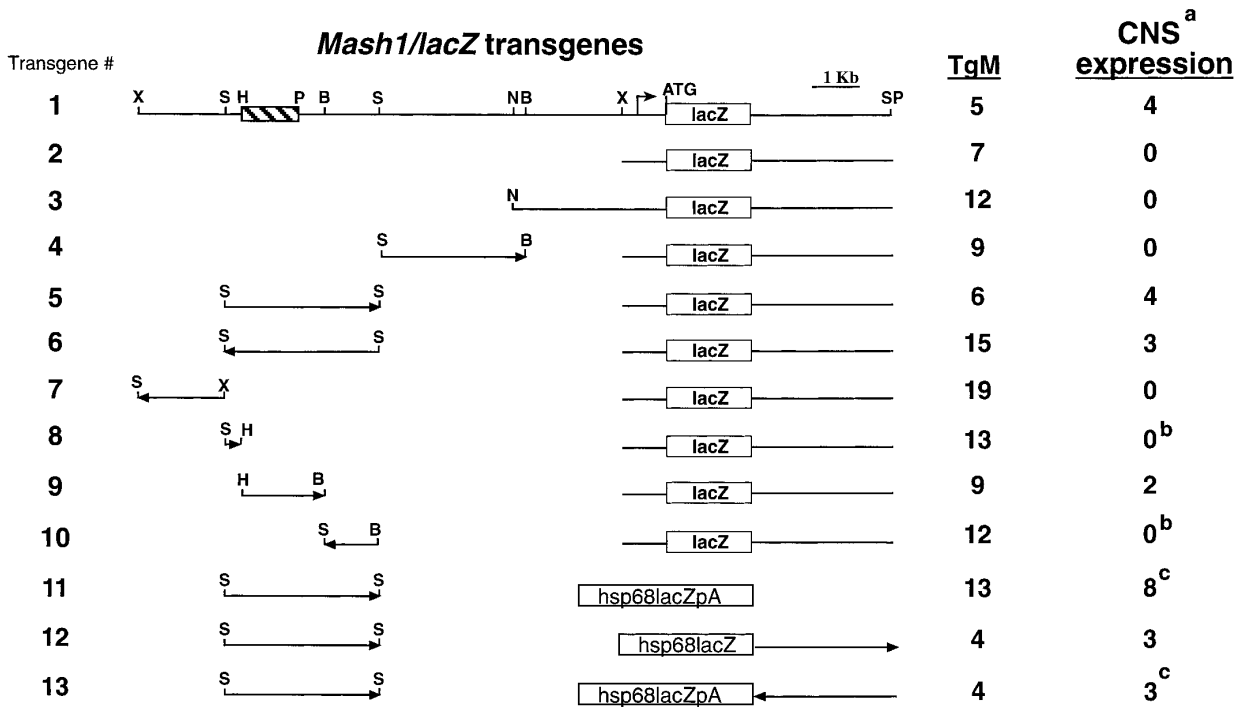

FIG. 1. Localization of an enhancer 5′ of the *Mash1* coding sequence. CNS expression of *Mash1/lacZ* reporter transgenes in E11.5 transgenic embryos. Transgenes 1–10 contain the *lacZ* coding sequence flanked by *Mash1* sequences on the 5′ and 3′ ends. Specifically, the 5′ sequence contains 315 bp upstream of the transcription start site and the entire untranslated region (575 bp). The 3′ sequence contains 2.7-kb *Mash1* flanking sequence including the 3′ untranslated region and intron found there. Transgenes 11–13 use a *lacZ* driven by the *hsp68* basal promoter. The poly(A) adenylation signal in Nos. 11 and 13 is contributed by SV40 and in No. 12 by *Mash1* sequences. All embryos were analyzed for *lacZ* expression as founder embryos except for transgene 1 embryos which come from established lines. Data are presented as the number of independent transgenic embryos expressing the transgene in the *Mash1*-specific CNS pattern (CNS expression) per the number of transgenic embryos identified by DNA analysis (TgM). Lines indicate the *Mash1* genomic sequence and the arrows indicate the orientation of the fragments. The location of the transcription start site (bent arrow) and the translation start (ATG) are shown in transgene 1. The hatched box indicates the location of the 5′ enhancer. (a) CNS expression includes the ventral telencephalon, thalamic and hypothalamic regions of the diencephalon, dorsal mesencephalon, metencephalon, myelencephalon, and the dorsal spinal neural tube (see Fig. 2B for example). (b) 1 of 13 (transgene 8) and 1 of 12 (transgene 10) embryos had low-level expression in diencephalon. (c) Expression in the ventral telencephalon, the thalamic region of the diencephalon, and the metencephalon is lacking in these embryos. B, *Bam*HI; H, *Hpa*I; N, *Nco*I; P, *Pst*I; S, *Sac*I; SP, *Sph*I; X, *Xba*I.

Mash1 coding region, when coupled with the *Mash1* basal reporter consistently directs *lacZ* expression to *Mash1*-expressing regions in the CNS (Fig. 1, transgene 5; Fig. 2B). Fragments outside this 3.3-kb sequence fail to drive *lacZ* expression in these regions (Fig. 1; transgenes 3, 4, and 7). The 3.3-kb fragment functions in a *Mash1*-specific manner in the CNS irrespective of its orientation (Fig. 1; transgene 6), although in the orientation opposite to wild type, it functions less efficiently. The 3.3-kb *Sac*I fragment can be replaced, although much less efficiently, by a central 1.6-kb *Hpa*I/*Bam*HI fragment (Fig. 1, transgene 9). Other sequences outside of the 1.6 kb but within the 3.3 kb are unable to efficiently drive expression of the transgene in the appropriate CNS regions (Fig. 1, transgenes 8 and 10). Taken together these results suggest a transcriptional enhancer sufficient for driving *Mash1* expression in the CNS is contained within a 1.6-kb sequence located approximately 7 kb 5′ of the *Mash1* coding region. This does not exclude the possibility that additional regulatory elements are present elsewhere in the 3.3-kb sequence.

A Role for Mash1 3′ Sequence for Efficient Expression in Multiple CNS Regions

The distal 5′ 3.3-kb *Sac*I fragment was tested for enhancer activity in combination with a heterologous promoter to test whether it could function efficiently without the basal *Mash1* promoter or 3′ sequences. We used the reporter *hsp68lacZpA* (Kothary *et al.*, 1989) which contains the *hsp68* basal promoter upstream of a *lacZ*-coding region followed by an SV40 polyadenylation signal cassette. The 3.3-kb *Sac*I fragment is sufficient to direct *lacZ* expression from the heterologous promoter but only in a subset of *Mash1*-expressing regions in the CNS (Fig. 1, transgene 11; Fig. 2C).

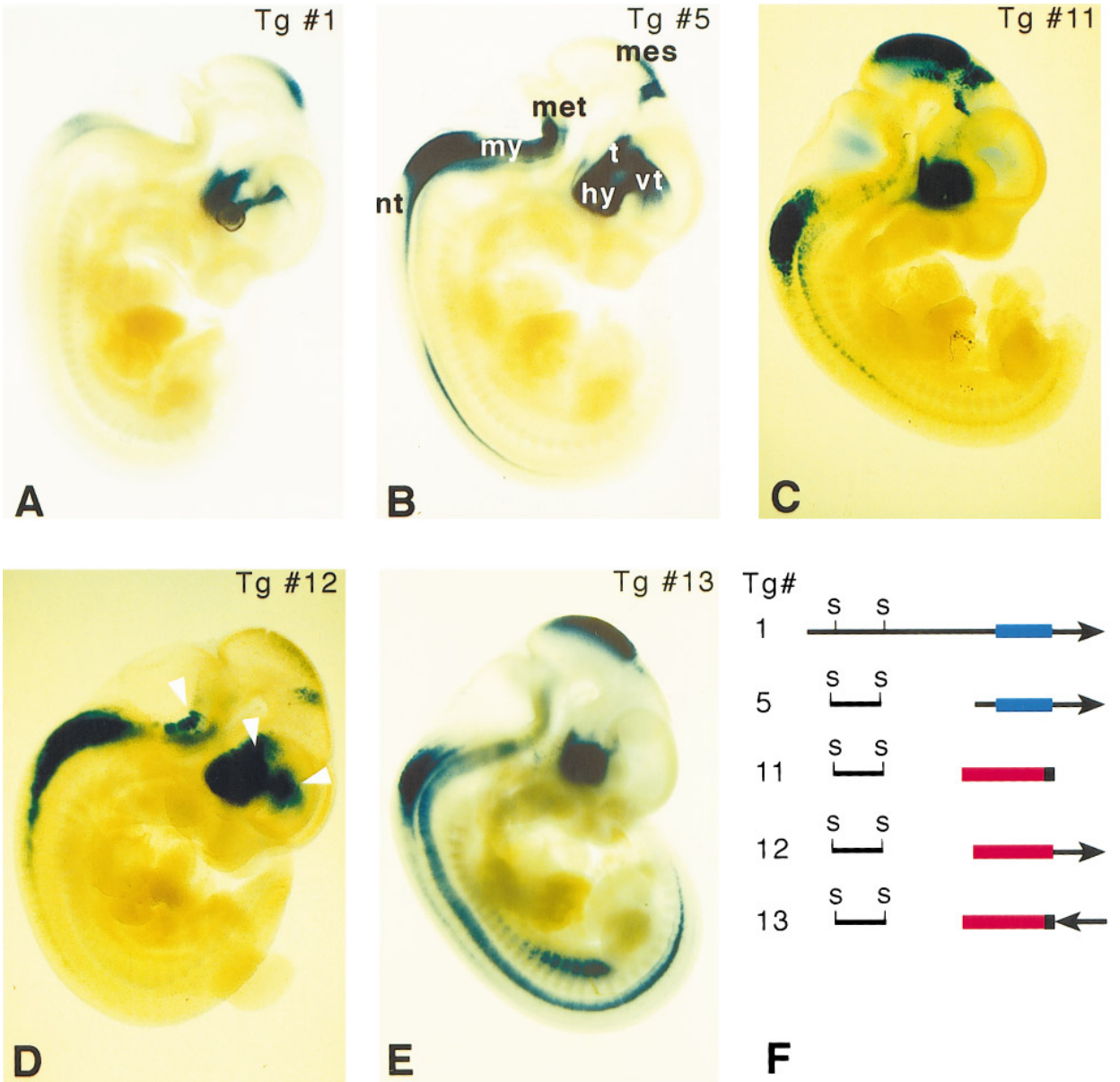

FIG. 2. *β*-Gal staining in *Mash1/lacZ* transgenic embryos. (A–E) Representative transgenic embryos at E11.5 stained for *β*-gal activity as whole mounts, cleared in methyl salicylate, and photographed under brightfield illumination. The transgene number is as indicated in each panel and refers to transgenes diagrammed in Fig. 1 and shown schematically in F. hy, hypothalamic region of diencephalon; mes, mesencephalon; met, metencephalon; my, myelencephalon; nt, dorsal spinal neural tube; t, thalamic region of the diencephalon; vt, ventral telencephalon. Arrowheads in D indicate the three regions of CNS requiring the *Mash1* 3′ sequence for expression. In F, the blue box is the *lacZ* coding region flanked by *Mash1* sequences (black lines). The red box is the *hsp68lacZ* reporter with the black box indicating the SV40 3′ cassette.

Consistent expression is observed only in the hypothalamic region of the diencephalon, dorsal mesencephalon, myelencephalon, and the dorsal spinal neural tube, but not in other *Mash1*-specific CNS regions including the ventral telencephalon, thalamic region of the diencephalon, and the metencephalon (compare Figs. 2B and 2C).

The loss of *lacZ* expression in specific CNS regions when the enhancer is combined with a heterologous promoter suggests that elements within the 0.9-kb 5′ or 2.7-kb 3′ flanking the *Mash1* coding region may contribute to expression in multiple regions of the CNS. However, these *Mash1* regions are not sufficient to drive reporter expression on

their own (Fig. 1, transgene 2). To determine if the 3′ sequence is responsible for the apparent regulatory activity, the SV40 polyadenylation signal cassette in the *Mash1*-enhancer/*hsp68lacZpA* transgene (No. 11) was replaced with the 2.7-kb *Mash1* 3′ sequence. In embryos carrying this reporter (No. 12), *LacZ* expression is recovered in all *Mash1*-specific CNS regions including the ventral telencephalon, thalamic region of the diencephalon, and the metencephalon (Fig. 1, transgene 12; Fig. 2D, arrowheads). Thus, the 3′ 2.7-kb sequence, which includes 3′ untranslated sequence, appears to cooperate with the distal enhancer to yield detectable levels of transgene expression in these three *Mash1*-expressing CNS domains.

Two possibilities exist for the function of the 3′ sequence in regulating *Mash1* expression in the CNS. First, the 2.7-kb sequence may contain enhancer elements that direct expression only when combined with additional elements in the 5′ enhancer. Alternatively, since the 3′ *Mash1* sequence also contributes the 3′ UTR, it is possible that post-transcriptional regulation occurs, and transcripts made with the *Mash1* 3′ sequences are translated more efficiently, or are more stable, than those containing the SV40 sequence. To help distinguish between these possibilities, we tested the 3′ sequence but in the opposite orientation (transgene 13). Thus, this transgene uses the poly(A) addition cassette from SV40 instead of from *Mash1,* but would still allow enhancer elements within the 2.7-kb 3′ sequence to be revealed. In embryos containing this transgene, expression was detected only in the partial CNS pattern and lacked expression in the ventral telencephalon, thalamic region of the diencephalon, and the myelencephalon (Fig. 1, transgene 13; Fig. 2E). These data suggest that the 3′ flanking sequence does not contain enhancer elements for *Mash1* expression, but rather the 3′ UTR within this sequence may harbor elements responsible for posttranscriptional regulation in specific regions of the CNS.

Multiple Regulatory Elements Are Contained within the Distal 5′ Enhancer

We observed that as the size of the enhancer decreased, the efficiency of obtaining transgenic embryos with high levels of expression also decreased. We also observed that if we tested the enhancer with the *hsp68lacZpA* reporter, we consistently obtained a high level of expression in a subset of CNS regions as outlined above. The smallest *Mash1* enhancer fragment that functions efficiently in this transgenic assay is a 1158-bp *Hpa*I/*Pst*I fragment found within the 3.3-kb enhancer (hatched box in Fig. 1, transgene 1). This 1158-bp enhancer is sufficient to drive *Mash1*-specific expression to the CNS in 5 of 6 embryos assayed (Fig. 3, transgene 14; Fig. 4A). Although sequence analysis of the 1158-bp enhancer revealed sequences resembling transcription factor consensus binding sites for homeodomain factors and bHLH factors, it was uninformative in revealing any specific regulatory components.

A series of deletions of this 1158-bp enhancer were tested to further define the sequence sufficient to regulate *Mash1*-specific CNS expression. The effect of each deletion in transgenic embryos was evaluated by the percentage of transgenic embryos expressing *lacZ* in the CNS, the loss of expression in specific CNS regions, and the overall expression levels when compared to the full-length 1158-bp enhancer (transgene 14). Each deletion was tested on the *hsp68lacZpA* heterologous reporter and assayed in transgenic embryos at E11.5.

Deletions of the 1158-bp sequence either on the 5′ and/or the 3′ end were tested to identify boundaries of the enhancer. Deletions of 177 bp from the 5′ end or 118 bp from the 3′ end of the enhancer were tolerated and expression was detected in the CNS regions seen with the full-length enhancer (Fig. 3, transgenes 15 and 19; Figs. 4B and 4C). However, when these sequences were deleted in combination, there was a significant loss of enhancer activity (compare Fig. 3, transgenes 15 and 19 with 22; Figs. 4B and 4C with 4D). These results suggest that activator elements are contained within nucleotides 1–177 and 1040–1158, and when both elements are deleted only inefficient expression of the transgene is maintained by elements which are present within the internal nucleotides 178–1039. Additional deletions from either the 5′ and/or 3′ ends resulted in significant loss of enhancer activity (Fig. 3, transgenes 16–26). This 5′ and 3′ enhancer deletion analysis suggests that there are multiple elements within the 1158-bp sequence required for enhancer activity. Further characterization of these elements is provided below.

Deletion of a Putative Negative Regulatory Region Reveals Additional Properties of the Enhancer

The presence of a negative element(s) is suggested with the deletion of 118 bp from the 3′ end of the 1158-bp enhancer (transgene 15). This deletion results in an overall increase in expression from the transgene in *Mash1*-specific regions (compare Figs. 4A and 4B). Strikingly, in 4 of the 7 embryos, β-gal activity was detected in the complete CNS pattern, the pattern seen previously only when the *Mash1* 3′ flanking sequence was present in the transgene. Thus, deletion of 118 bp on the 3′ end of the enhancer results in expression not only in the hypothalamic region of the diencephalon, dorsal mesencephalon, myelencephalon, and the dorsal spinal neural tube, but also in the ventral telencephalon, thalamic region of the diencephalon, and the metencephalon (Fig. 4B, arrowheads). These data suggest that a negative regulatory element(s) is contained within the 118-bp sequence between nucleotides 1040 and 1158 of the enhancer since when deleted, both an increased overall expression level and expression in additional *Mash1*-specific regions is observed. In addition, these data demonstrate the presence of positive elements for the complete CNS pattern within nucleotides 1–1039 of the enhancer that function without additional *Mash1* 3′ sequence.

Surprisingly, in several transgenic embryos containing nucleotides 1–1039, expression was observed in regions

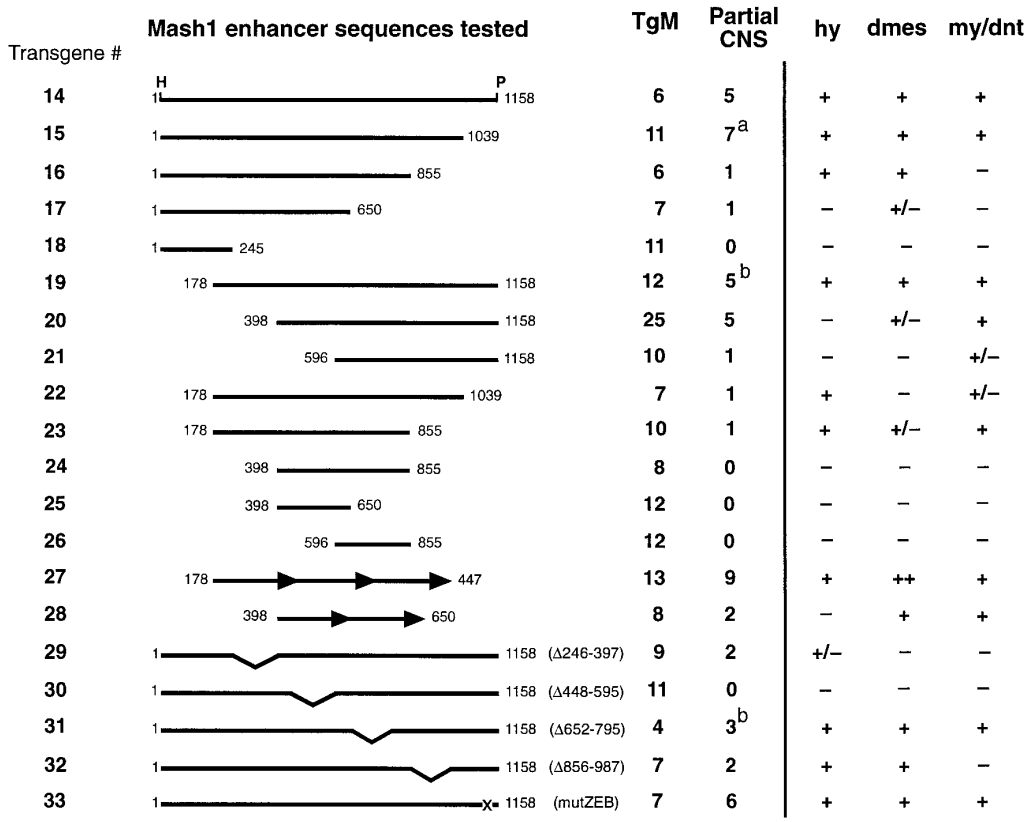

FIG. 3. Multiple regulatory elements within the 5′ enhancer. All transgenes test the indicated enhancer sequences in combination with the *hsp68lacZpA* reporter in transgenic mice at E11.5. Data are presented as the total number of transgenic embryos identified by DNA analysis (TgM), and the number of embryos expressing *lacZ* within a subset of *Mash1*-specific regions in the CNS (partial CNS). Within the CNS, *lacZ* expression is scored in the hypothalamic region of the diencephalon (hy), dorsal mesencephalon (dmes), and the myelencephalon and/or the dorsal spinal neural tube (my/dnt). + indicates easily detectable, +/− indicates barely detectable and/or expression in a low percentage of transgenic embryos, and − indicates lack of *lacZ* expression in that particular CNS region. Refer to Fig. 1 (hatched box) for the location of the enhancer relative to the *Mash1* coding region. (a) In multiple embryos with transgene 15, expression is also observed in the ventral telencephalon, thalamic region of the diencephalon, the metencephalon, and in olfactory epithelium and enteric regions. (b) In 1 of 5 embryos (transgene 19), and 1 of 3 embryos (transgene 31), expression is observed in the ventral telencephalon.

containing the sympathetic, enteric, and olfactory neuronal precursors (Fig. 4B, arrows). Expression in these regions accurately reflects the endogenous *Mash1* expression pattern. Expression in the sympathetic and enteric precursors was previously detected in *Mash1/lacZ* transgenic embryos only when 36 kb of *Mash1*-flanking sequence was included (Verma-Kurvari *et al.*, 1996). Expression of *Mash1/lacZ* transgenes in olfactory epithelium has never been detected even with the largest transgenic constructs (Verma-Kurvari *et al.*, 1996). While the expression of *lacZ* in the sympathetic precursors was detected in only one transgenic embryo containing the 118-bp deleted enhancer, expression in the olfactory epithelium and enteric regions was observed in several embryos (Fig. 4B and data not shown). Thus, deletion of a putative negative element may be revealing positive elements for expression in the PNS and olfactory regions, not previously detected, which colocalize with the CNS elements.

The sequence of the 118 bp containing the putative negative element contains a consensus binding site for the ZEB Zn finger transcription factor. *In vitro* studies have demonstrated that ZEB binds to an extended E-box sequence (GCAGGTG) within an enhancer of the immunoglobulin heavy chain gene where it appears to act as a repressor of transcription (Genetta *et al.*, 1994). To test directly whether this sequence was responsible for the repressor activity observed in the 118-bp deletion (transgene 15), we mutated the ZEB consensus binding site to (CTGTTCA), which also mutates the overlapping E-box (CAGGTG), and tested its function in transgenic embryos. Embryos expressing the mutated transgene show higher levels of expression in general, but none give the complete CNS/PNS *Mash1* pattern as seen with the 118-bp deletion (Fig. 3, transgene 33; Fig. 4E). It appears that the ZEB/E-box site contributes to the level of activity of the 1158-bp enhancer but additional elements within the 118-bp region may be required for full inhibitory activity.

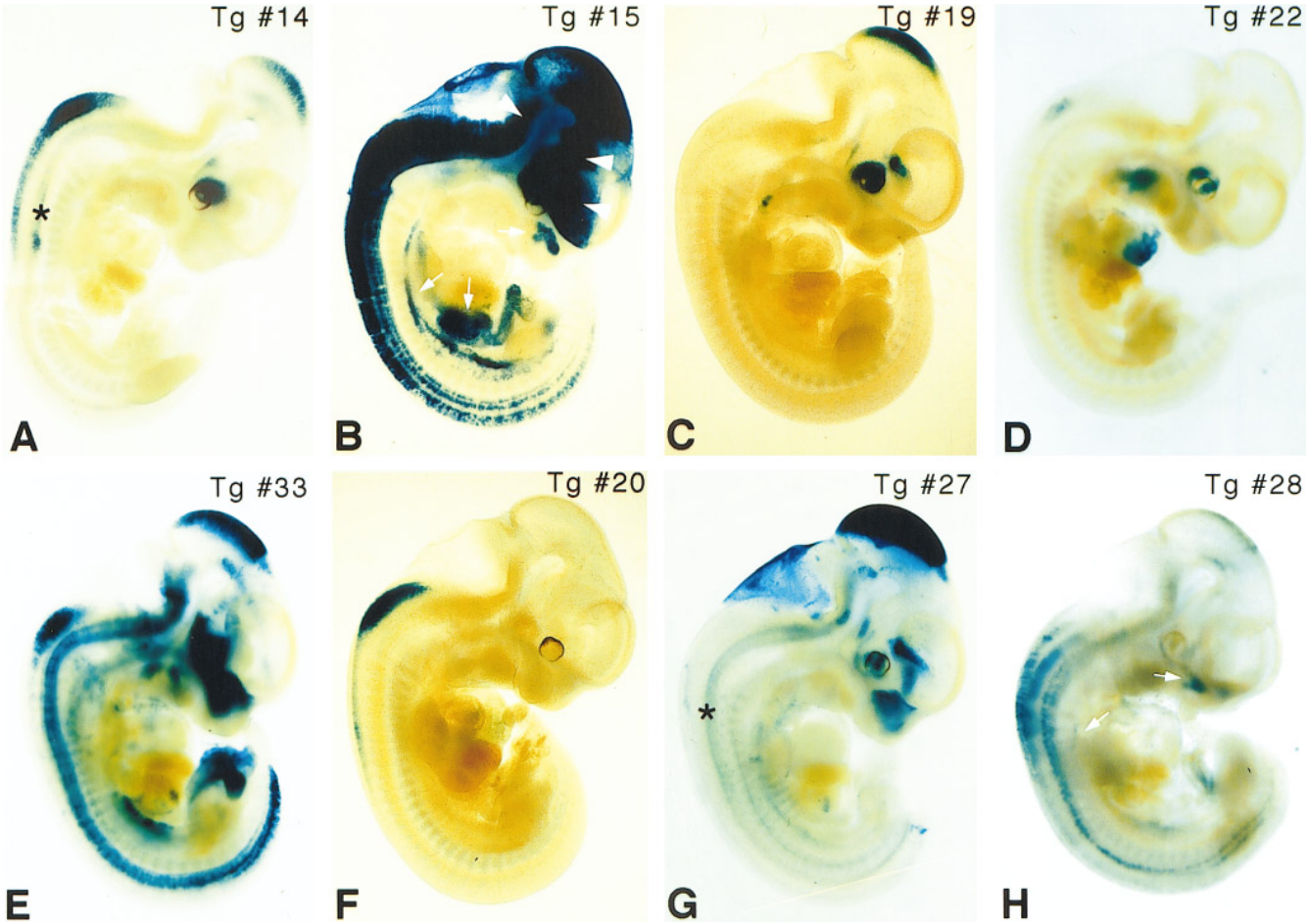

FIG. 4. β-Gal staining in *Mash1/hsp68lacZpA* transgenic embryos. (A–H) Representative transgenic embryos at E11.5 stained for β-gal activity as whole mounts, cleared in methyl salicylate, and photographed under brightfield illumination. The transgene number is as indicated in each panel and refers to transgenes diagrammed in Fig. 3. Expression in the ventral telencephalon, thalamic region of the diencephalon, and the metencephalon is indicated by arrowheads in B. *LacZ* expression in *Mash1*-specific regions of the PNS and the olfactory epithelium is indicated by arrows in B and H. Ectopic expression (*) in ventral neural tube is commonly observed in embryos injected with *Mash1/hsp68lacZpA* constructs.

Identification of Region-Specific Elements within the Enhancer

An interesting question is whether within the enhancer, different elements function to drive transcription in specific regions within the *Mash1*-specific expression domains. For example, is there an element that is required for expression in the hypothalamic region of the diencephalon versus the dorsal mesencephalon or dorsal spinal neural tube? The inefficient expression from the enhancer deletions, visualized as the low percentage of transgenic embryos expressing, make this type of conclusion difficult in some cases such as in transgenes 16 and 23 where there is only a single expressing embryo out of 6 and 10 transgenic embryos, respectively. However, for transgene 20 which deletes 397 bp from the 5′ end, there are 5 expressing embryos all lacking expression in the hypothalamic region of the diencephalon and very minimal expression in the dorsal mesencephalon (Figs. 3, 4F). A transgene deleting 177 bp on the 5′ end does not effect expression in these regions (Fig. 3, transgene 19; compare Fig. 4C with 4F). Together, these data suggest elements important for expression in the hypothalamic region of the diencephalon and possibly the dorsal mesencephalon are contained within nucleotides 178–397.

To examine the function of this sequence in more detail, we tested whether it was necessary and/or sufficient for enhancer function. A deletion between nucleotides 246–397 severely compromises the ability of the enhancer to function not only in the diencephalon, but also in the other CNS regions (Fig. 3, transgene 29). Two of the three *lacZ*-expressing embryos with this deletion had extremely high ectopic expression throughout the nonneural tissues. Vibratome sec-

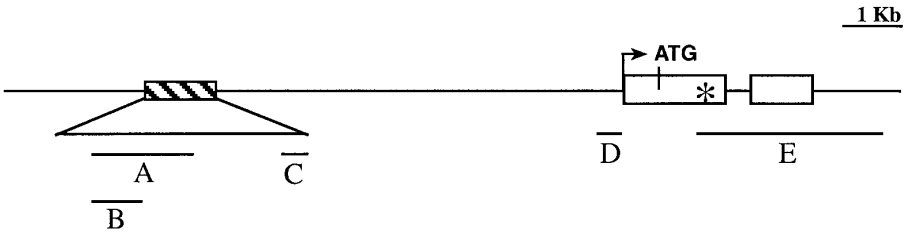

FIG. 5. Summary diagram of sequences regulating *Mash1* expression. 13 kb from the *Mash1* locus is shown. The open boxes represent the *Mash1*-transcribed region, ATG is the start codon, and the asterisk is the stop codon. The hatched box represents the 1158-bp 5′ *Mash1* enhancer. Regions functioning in regulating *Mash1* expression are indicated as A–E. (A) nucleotides 178–650, containing positive elements for expression in the dorsal mesencephalon, myelencephalon, and dorsal spinal neural tube; (B) nucleotides 178–447, containing positive elements for diencephalon expression. Deletions within regions A and B dramatically reduce the activity of the enhancer. C contains negative elements, some of which may be attributed to a ZEB/E-box site within the sequence. D is the 315-bp sequence that acts as a basal promoter in transgenes 1–10. Sequence in E contains regulatory information, possibly at a posttranscriptional level, for efficient expression of the transgene in the ventral telencephalon, the thalamic region of the diencephalon, and the metencephalon, when combined with the 5′ enhancer.

tions of these embryos revealed some expression in the hypothalamic region of the diencephalon but no expression in other *Mash1*-specific CNS regions. These results combined with the 5′ deletions suggest that nucleotides 246–397 harbor essential positive elements for efficient *Mash1*-specific CNS expression. We next addressed the question of whether this sequence is sufficient to drive expression to any of the *Mash1*-specific CNS regions. Data from transgene 17, which tests nucleotides 1–650, suggest that as a single copy this region is not sufficient to drive expression. However, since regulatory elements often do not function efficiently in these types of assays as a single copy, we tested the sequence from nucleotides 178 to 447 as a tandem array of 3 copies for enhancer activity. Consistent strong expression from this transgene was observed in the dorsal mesencephalon with variable expression in the hypothalamic region of the diencephalon and in the myelencephalon/dorsal spinal neural tube (Fig. 3, transgene 27; Fig. 4G). Expression was also consistently observed in regions around the ventral telencephalon and the olfactory epithelium. However, in both cases, the expression does not reflect the correct *Mash1* pattern. These data demonstrate the presence of a regulatory element(s) within nucleotides 178–447 sufficient to drive transcription specifically in a subset of the *Mash1* CNS pattern most strongly in the dorsal mesencephalon.

A second region was identified that also appears to have an essential role in the function of the 1158-bp enhancer. A deletion of nucleotides 448–595 within the context of the enhancer completely disrupts enhancer activity (Fig. 3, transgene 30). Disruption of enhancer function with internal deletions such as this suggests that an essential element lies within the deletion or that spacing between remaining regulatory elements is disrupted. To determine whether this region is sufficient to drive *Mash1*-specific expression, one and two copies of the sequence from 398 to 650 were tested. One copy of this sequence is not sufficient to drive expression (Fig. 3, transgene 25). However, when two copies of this sequence are present in tandem, *Mash1*-specific expression was detected at a low frequency in the dorsal mesencephalon and the dorsal spinal neural tube (Fig. 3, transgene 28; Fig. 4H). Surprisingly, in one of these embryos, expression was also detected in the olfactory epithelium and the sympathetic region (Fig. 4H, arrows). We conclude that sequence within nucleotides 398–650 is essential for full enhancer function, and that it contains an element(s) able to drive transcription specifically to the dorsal mesencephalon and dorsal spinal neural tube.

Two additional deletions within the enhancer were tested for their ability to disrupt enhancer function. Deletion of nucleotides 652–795 or 856–987 within the context of the 1158-bp enhancer did not have dramatic consequences for enhancer function (Fig. 3, transgenes 31 and 32).

DISCUSSION

Little is known about the factors that control the expression of essential neural regulatory genes of the bHLH family of transcription factors. Expression of these genes is tightly regulated both spatially and temporally in the developing nervous system, suggesting that this regulation is important for proper development to occur. Previous studies of *Mash1* regulation suggested regulatory elements governing different domains of expression are distributed broadly in greater than 36 kb flanking the *Mash1* coding region. Here we demonstrate the complexity in the regulation of *Mash1* at multiple levels. We localize a 1158-bp CNS enhancer to a region ~7 kb 5′ of the coding sequence. This enhancer contains multiple positive and negative elements that appear to act together to yield the *Mash1* expression pattern in specific regions of the developing CNS (see summary diagram, Fig. 5).

The clustering of positive and negative *cis*-acting regulatory elements within a restricted region of DNA demonstrated here for *Mash1* has been described previously for

many genes in sea urchin, *Drosophila,* and mammals (Kruse *et al.,* 1993; Li *et al.,* 1996; Morrison *et al.,* 1996; Song *et al.,* 1996). In genes that are regulated by multiple positive elements, such as *HOXD4,* often a single regulatory element by itself is not sufficient for expression; however, deletion of the same element is enough to abolish expression (Morrison *et al.,* 1996). This appears to be the case for elements within the 1158-bp *Mash1* enhancer as well. Only when the sequences are tested as multimers is specific expression revealed.

Genes that show complex expression patterns in multiple cell types often have separate elements for expression in the different regions (Small *et al.,* 1992; Kruse *et al.,* 1993; Logan *et al.,* 1993; Zimmerman *et al.,* 1994; Li *et al.,* 1996). Previous work testing 36 kb of *Mash1*-flanking sequence suggested that elements for *Mash1* expression in the PNS, or the olfactory epithelium and the retina, are distinct from each other and from those functioning in CNS expression (Verma-Kurvari *et al.,* 1996). Here we have identified an enhancer that functions efficiently for CNS expression. However, data from multiple transgenes suggest that positive elements for expression in the PNS and olfactory epithelium colocalize with this CNS enhancer. The transgene containing nucleotides 1–1039 (transgene 15) functioned reproducibly to drive expression in the PNS and olfactory epithelium (Fig. 4B). In addition, in one embryo, a multimer of nucleotides 397–650 drove expression to these regions (Fig. 4H). Thus, it seems likely that PNS- and olfactory epithelium-specific elements reside within the CNS enhancer characterized here but they are inefficient, or more effectively repressed within the context of the larger transgene.

Posttranscriptional regulation is another aspect of *Mash1* regulation suggested by these studies. Sequence 3′ of the *Mash1* coding sequence when combined with the 5′ enhancer directs *lacZ* expression in the ventral telencephalon, the thalamic region of the diencephalon, and the metencephalon. Expression in these regions is lost in transgenes that lack the 3′ *Mash1* sequence. These observations can not be simply explained by an enhancer specific for these regions of the CNS since the 3′ sequence alone has no visible activity, and the 3′ sequence does not function when present in the opposite orientation. The 3′ *Mash1* sequence contributes both transcribed and untranscribed regions to the transgene. Sequences contained in 3′ UTRs have been shown to harbor elements responsible for posttranscriptional regulation of many genes in a variety of organisms (Wickens *et al.,* 1997). It is possible that the differences in the 3′ UTR are responsible for differential regulation of *lacZ* mRNA in the CNS such that transcripts containing *Mash1* 3′ UTR are more stable, processed more efficiently, or translated more efficiently than those containing the SV40 sequence.

Conservation of Regulatory Mechanisms between Drosophila and Mammals

Mash1 is a mammalian homolog of the *Drosophila achaete–scute* genes that shows conservation at both the level of sequence and function. Therefore, it is of interest to consider whether there is also conservation of regulatory mechanisms. Genetic and biochemical studies in *Drosophila* have identified many candidate genes which regulate *achaete–scute* expression (Skeath *et al.,* 1992; Ramain *et al.,* 1993; Ohsako *et al.,* 1994; van Doren *et al.,* 1994; Jiminez *et al.,* 1995; Cubadda *et al.,* 1997). The two homeodomain proteins, ARA and CAUP, upregulate transcription of *achaete* and *scute* genes and ARA has been shown to interact directly with the sequences in the *achaete–scute* enhancer (Gómez-Skarmeta *et al.,* 1996). Sequence analysis of the 1158-bp *Mash1* enhancer indicates the presence of several A/T-rich regions which include multiple consensus binding sites for homeodomain proteins. The specific binding site for ARA, as identified in the *achaete–scute* enhancer, however, does not have a match within the *Mash1* enhancer sequence and the binding site for CAUP has not been characterized. Mammalian homologs of *ara* and *caup* have been identified (P. Gruss, personal communication). Whether these mammalian homologs regulate *Mash1* expression awaits characterization of their expression pattern, the ability to bind to *Mash1* enhancer sequences, and the effect of mutating their binding sites.

ACKNOWLEDGMENTS

We gratefully acknowledge Dr. Raymond MacDonald for critical comments on the manuscript, members of the Johnson laboratory for ongoing discussions of this work, and Shirley Hall in the Micro Sequencing Facility at UTSW for providing quality sequence data. J.E.J. is an Established Investigator for the American Heart Association. This work was funded by NIH RO1-NS32817, the Perot Foundation, and the Kent Waldrep National Paralysis Foundation.

REFERENCES

Ahmad, I. (1995). Mash-1 is expressed during ROD photoreceptor differentiation and binds to an E-box $E_{opsin\text{-}1}$ in the rat opsin gene. *Dev. Brain Res.* **90,** 184–189.

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.

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

Begley, C. G., Lipkowitz, S., Gobel, V., Mahon, K. A., Bertness, V., Green, A. R., Gough, N. M., and Kirsch, I. R. (1992). Molecular characterization of NSCL, a gene encoding a helix–loop–helix protein expressed in the developing nervous system. *Proc. Natl. Acad. Sci. USA* **89,** 38–42.

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. *Hum. Mol. Genet.* **5,** 1207–1216.

Campuzano, S., and Modolell, J. (1992). Patterning of the *Drosophila* nervous system: The *achaete–scute* gene complex. *Trends Genet.* **8,** 202–208.

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.

Ferreiro, B., Kintner, C., Zimmerman, K., Anderson, D., and Harris, W. A. (1994). XASH genes promote neurogenesis in *Xenopus* embryos. *Development* **120,** 3649–3655.

Genetta, T., Ruezinsky, D., and Kadesch, T. (1994). Displacement of an E-box-binding repressor by basic helix–loop–helix proteins: Implications for B-cell specificity of the immunoglobulin heavy-chain enhancer. *Mol. Cell. Biol.* **14,** 6153–6163.

Gómez-Skarmeta, J. L., Diez del Corral, R., de la Calle-Mustienes, E., Ferrés-Marcó, 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.

Gordon, M. K., Mumm, J. S., Davis, R. A., Holcomb, J. D., and Calof, A. L. (1995). Dynamics of MASH1 expression *in vitro* and *in vivo* suggest a non-stem cell site of MASH1 action in the olfactory receptor neuron lineage. *Mol. Cell. Neurosci.* **6,** 363–379.

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.

Guillemot, F., and Joyner, A. (1993). Expression of murine Achaete–Scute and Notch homologues in the developing central nervous system. *Mech. Dev.* **42,** 171–185.

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.

Helms, A. W., and Johnson, J. E. (1998). Progenitors of dorsal commissural interneurons are defined by MATH1 expression. *Development* **125,** 919–925.

Ishibashi, M., Ang, S.-L., Shiota, K., Nakanishi, S., Kageyama, R., and Guillemot, F. (1995). Targeted disruption of mammalian *hairy* and *Enhancer of split* homolog-1 (*HES-1*) leads to up-regulation of neural helix–loop–helix factors, premature neurogenesis, and severe neural tube defects. *Genes Dev.* **9,** 3136–3148.

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.

Jasoni, C. L., and Reh, T. L. (1996). Temporal and spatial pattern of MASH-1 expression in the developing rat retina demonstrates progenitor cell heterogeneity. *J. Comp. Neurol.* **369,** 319–327.

Jimenez, F., Martin-Morris, L. E., Velasco, L., Chu, H., Sierra, J., Rosen, D. R., and White, K. (1995). vnd, a gene required for early neurogenesis of *Drosophila*, encodes a homeodomain protein. *EMBO J* **14,** 3487–3495.

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. *Crit. Rev. Neurobiol.* **9,** 177–188.

Kawakami, H., Maruyama, H., Yasunami, M., Ohkubo, H., Hara, H., Saida, T., Nakanishi, S., and Nakamura, S. (1996). Cloning and expression of a rat brain basic helix–loop–helix factor. *Biochem. Biophys. Res. Commun.* **221,** 199–204.

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.

Kothary, R., Clapoff, S., Darling, S., Perry, M. D., Moran, L. A., and Rossant, J. (1989). Inducible expression of an hsp68-lacZ hybrid gene in transgenic mice. *Development* **105,** 707–714.

Kruse, F., Rose, S. D., Swift, G. H., Hammer, R. E., and MacDonald, R. J. (1993). An endocrine-specific element is an integral component of an exocrine- specific pancreatic enhancer. *Gene Dev.* **7,** 774–786.

Kume, H., Maruyama, K., Tomita, T., Iwatsubo, T., Saido, T. C., and Obata, K. (1996). Molecular cloning of a novel basic helix–loop–helix protein from the rat brain. *Biochem. Biophys. Res. Commun.* **219,** 526–530.

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.

Lemercier, C., To, R. Q., Swanson, B. J., Lyons, G. E., and Konieczny, S. F. (1997). Mist1: A novel basic helix–loop–helix transcription factor exhibits a developmentally regulated expression pattern. *Dev. Biol.* **182,** 101–113.

Li, L., Miano, J. M., Mercer, B., and Olson, E. N. (1996). Expression of the SM22α promoter in transgenic mice provides evidence for distinct transcriptional regulatory programs in vascular and visceral smooth muscle cells. *J. Cell Biol.* **132,** 849–859.

Lo, L., and Anderson, D. J. (1995). Postmigratory neural crest cells expressing c-RET display restricted developmental and proliferative capacities. *Neuron* **15,** 527–539.

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

Logan, C., Khoo, W. K., Cado, D., and Joyner, A. L. (1993). Two enhancer regions in the mouse EN-2 locus direct expression to the mid/hindbrain region and mandibular myoblasts. *Development* **117,** 905–916.

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

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. *Mol. Cell. Biol.* **16,** 5792–5800.

Morrison, A., Moroni, M. C., Ariza-McNaughton, L., Krumlauf, R., and Mavilio, F. (1996). In vitro and transgenic analysis of a human HOXD4 retinoid-responsive enhancer. *Development* **122,** 1895–1907.

Ohsako, S., Hyer, J., Panganiban, G., Oliver, I., and Caudy, M. (1994). hairy function as a DNA-binding helix–loop–helix repressor of *Drosophila* sensory organ formation. *Genes Dev.* **8,** 2743–2755.

Porteus, M. H., Bulfone, A., Liu, J.-K., Puelles, L., Lo, L.-C., and Rubenstein, J. L. R. (1994). DLX-2, MASH-1, and MAP-2 expression and bromodeoxyuridine incorporation define molecularly distinct cell populations in the embryonic mouse forebrain. *J. Neurosci.* **14,** 6370–6383.

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.

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., Panganiban, G., Selegue, J., and Carroll, S. B. (1992). Gene regulation in two dimensions: The proneural achaete and scute genes are controlled by combinations of axis-patterning genes through a common intergenic control region. *Genes Dev.* **6,** 2606–2619.

Small, S., Blair, A., and Levine, M. (1992). Regulation of even-skipped stripe 2 in the *Drosophila* embryo. *EMBO J.* **11,** 4047–4057.

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. *Mol. Cell. Neurosci.* **8,** 221–241.

Sommer, L., Shah, N., Rao, M., and Anderson, D. J. (1995). The cellular function of MASH1 in autonomic neurogenesis. *Neuron* **15,** 1245–1258.

Song, D. L., Chalepakis, G., Gruss, P., and Joyner, A. L. (1996). Two Pax-binding sites are required for early embryonic brain expression of an Engrailed-2 transgene. *Development* **122,** 627–635.

Takebayashi, K., Takahashi, S., Yokota, C., Tsuda, H., Nakanishi, S., Asashima, M., and Kageyama, R. (1997). Conversion of ectoderm into a neural plate by ATH-3, a vertebrate basic helix–loop–helix gene homologous to *Drosophila* proneural gene atonal. *EMBO J.* **16,** 384–395.

Turner, D. L., and Weintraub, H. (1994). Expression of achaete–scute homolog 3 in *Xenopus* embryos converts ectodermal cells to a neural fate. *Genes Dev.* **8,** 1434–1447.

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.

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.

Wickens, M., Anderson, P., and Jackson, R. J. (1997). Life and death in the cytoplasm: Messages from the 3′ end. *Curr. Opin. Genet. Dev.* **7,** 220–232.

Yasunami, M., Suzuki, K., Maruyama, H., Kawakami, H., Nagai, Y., Hagiwara, M., and Ohkubo, H. (1996). Molecular cloning and characterization of a cDNA encoding a novel basic helix–loop–helix protein structurally related to NeuroD/BHF1. *Biochem. Biophys. Res. Commun.* **220,** 754–758.

Yun, K., and Wold, B. (1996). Skeletal muscle determination and differentiation: story of a core regulatory network and its context. *Curr. Opin. Cell Biol.* **8,** 877–889.

Zimmerman, L., Lendahl, U., Cunningham, M., McKay, R., Parr, B., Gavin, B., Mann, J., Vassileva, G., and McMahon, A. (1994). Independent regulatory elements in the nestin gene direct transgene expression to neural stem cells or muscle precursors. *Neuron* **12,** 11–24.

Received for publication June 27, 1997
Accepted February 3, 1998