Molecular and Cellular Neuroscience 7, 1–16 (1996)
Article No. 0001

MCN

Expression Patterns of Two Murine Homologs of *Drosophila Single-Minded* Suggest Possible Roles in Embryonic Patterning and in the Pathogenesis of Down Syndrome

Chen-Ming Fan,*[,1] Ellen Kuwana,*[,†] Alessandro Bulfone,† Colin F. Fletcher,‡ Neal G. Copeland,‡ Nancy A. Jenkins,‡ Stephen Crews,§ Salvador Martinez,¶ Luis Puelles,¶ John L. R. Rubenstein,† and Marc Tessier-Lavigne*

**Howard Hughes Medical Institute, Programs in Cell Biology, Developmental Biology and Neuroscience, Department of Anatomy, University of California, San Francisco, 513 Parnassus Avenue, San Francisco, California 94143-0452; †Nina Ireland Laboratory of Developmental Neurobiology, Programs in Developmental Biology and Neuroscience, Center for Neurobiology and Psychiatry, Department of Psychiatry and Langley Porter Psychiatric Institute, University of California, San Francisco, San Francisco, California 94143-0984; ‡Mammalian Genetics Laboratory, ABL-Basic Research Program, NCI-Frederick Cancer Research and Development Center, Frederick, Maryland 21702; §Department of Biochemistry and Biophysics, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599-7260; and ¶Department of Morphological Sciences, Faculty of Medicine, University of Murcia, 30100 Murcia, Spain*

The *single-minded* (*sim*) gene encodes a transcriptional regulator that functions as a key determinant of central nervous system (CNS) midline development in *Drosophila.* We report here the identification of two murine homologs of *sim*, *Sim1* and *Sim2*, whose products show a high degree of sequence conservation with *Drosophila* SIM in their amino-terminal halves, with each containing a basic helix–loop–helix domain as well as a PAS domain. *Sim1* maps to the proximal region of mouse chromosome 10, whereas *Sim2* maps to a portion of the distal end of chromosome 16 that is syntenic to the Down syndrome critical region of human chromosome 21. Recent exon-trapping studies have identified in the critical region several exons of a human *sim* homolog which appears to be the homolog of murine *Sim2;* this has led to the hypothesis that increased dosage of this *sim* homolog in cases of trisomy 21 might be a causal factor in the pathogenesis of Down syndrome. We have examined the expression patterns of the *Sim* genes during embryogenesis. Both genes are expressed in dynamic and selective fashion in specific neuromeric compartments of the developing forebrain, and the expression pattern of *Sim2* provides evidence for early regionalization of the diencephalon prior to any overt morphological differentiation in this region. Outside the CNS, *Sim1* is expressed in mesodermal and endodermal tissues, including developing somites, mesonephric duct, and foregut. *Sim2* is expressed in facial and trunk cartilage, as well as trunk muscles. Both murine *Sim* genes are also expressed in the developing kidney. Our data suggest that the *Sim* genes play roles in directing the regionalization of tissues where they are expressed. Moreover, the expression pattern documented for *Sim2* may provide insights into its potential roles in Down syndrome.

[1] Present address: Carnegie Institute of Washington, Department of Embryology, 115 West University Parkway, Baltimore, MD 21210.

INTRODUCTION

In *Drosophila*, the *single-minded* gene (*sim*) (Thomas *et al.*, 1988; Crews *et al.*, 1988) directs the development of cells located at the midline of the developing central nervous system (CNS) (reviewed by Crews *et al.*, 1992). In *sim* mutant embryos these cells fail to differentiate and eventually die (Thomas *et al.*, 1988). Conversely, when *sim* is expressed ubiquitously (under control of a heat-shock promoter) in transgenic flies, many other cells of

1044-7431/96 $18.00
Copyright © 1996 by Academic Press, Inc.
All rights of reproduction in any form reserved.

the ventral neurogenic region differentiate into CNS midline cells (Nambu *et al.*, 1991). Thus the *sim* gene product appears to be both necessary and sufficient for the differentiation of CNS midline cells within the ventral neurogenic region. In addition to its function in CNS midline cells, *sim* is expressed in a subset of myoblasts, the gut, and the brain, where it may also play roles in regionalization and cell fate determination (Crews *et al.*, 1988, Lewis and Crews, 1994).

The SIM protein contains a basic helix–loop–helix (bHLH) motif and localizes to the cell nucleus where it may act as a transcriptional activator (Nambu *et al.*, 1991; Franks and Crews, 1994). bHLH transcription factors appear to play important roles in directing cell fates and controlling cell proliferation and differentiation during embryonic development (see Weintraub *et al.*, 1991; Jan and Jan, 1993; Olson and Klein, 1994, for review). Members of this family include, for example, the products of the myogenic determination genes like *Myf5*, *MyoD*, and *Myogenin*, and the neurogenic genes of the *aechete–scute* complex. These proteins function by forming heterodimers through their helix–loop–helix (HLH) domains and bind to specific DNA sequences through their basic domains to transactivate target genes. Several of these genes are evolutionarily conserved in sequence, and in many cases homologs in widely divergent species from arthropods to vertebrates have been found to have apparently similar patterns of expression and embryonic functions (reviewed by Jan and Jan, 1993).

In addition to its bHLH domain, *Drosophila* SIM possesses a domain that shares homology with a domain in the *Drosophila* PER protein (which plays a role in controlling circadian rhythms) and the two subunits of the human aromatic hydrocarbon (dioxin) receptor, ARNT (the aromatic-hydrocarbon receptor nuclear translocator) and AHR (the aromatic-hydrocarbon receptor) (Crews *et al.*, 1988; Hoffman *et al.*, 1991; Burbach *et al.*, 1992). These domains have been termed PAS domains (for PER, ARNT, SIM). The PAS domain in PER has been shown to function as a dimerization domain (Huang *et al.*, 1993; Reisz-Porszasz *et al.*, 1994; Jain *et al.*, 1994; Lindebro *et al.*, 1995), whereas that of AHR functions as a ligand (dioxin)-binding and heat-shock protein association domain (Burbach *et al.*, 1992; Whitelaw *et al.*, 1993). The recently identified hypoxia-inducible factor 1 (HIF-1) also contains both bHLH and PAS domains and is a new member of this gene family (Wang *et al.*, 1995).

In searching for genes related to *sim* in the mammalian genome by low-stringency hybridization, we isolated two murine homologs of *sim*. Like the *Drosophila* SIM protein, the two predicted murine proteins each possess a bHLH and a PAS domain. To explore the possible functions of the mammalian *Sim* genes, we characterized their expression during mouse embryogenesis by *in situ* hybridization. Like *sim*, the mammalian homologs show restricted patterns of expression during embryonic development. *Sim1* and *Sim2* have restricted distributions in the developing brain, paraxial mesoderm, intermediate mesoderm, and gut that suggest their involvement in the regionalization of these tissues during early development. Importantly, during the final stages of this study, two studies reported the identification of fragments of a candidate human *SIM* gene by exon-trapping of DNA sequences within the Down syndrome critical region of human chromosome 21 (Chen *et al.*, 1995; Dahmane *et al.*, 1995). The open reading frame of the six trapped exons comprises a bHLH and a partial PAS domain highly homologous to those of the mouse *Sim2* gene. Our results describing the expression pattern of the murine homolog confirm and extend a report of the expression pattern of the *SIM2* gene in human and rat embryos (Dahmane *et al.*, 1995) and may provide insights to possible links between *SIM2* function and Down syndrome.

RESULTS

Isolation of cDNAs for Two Murine Homologues of Drosophila sim

To search for murine homologues of *Drosophila sim*, we screened a mouse genomic library at low stringency using a probe corresponding to the *sim* bHLH domain (Nambu *et al.*, 1991). Seven clones were isolated. The sequence of each hybridizing fragment contained a short open reading frame that was highly homologous over a 171-nt stretch to the sequence coding for the bHLH domain of *sim*. Five of the seven clones contained fragments of one gene, designated *Sim1*, and the other two contained fragments of a second gene, designated *Sim2*. Because *in situ* hybridization demonstrated the expression of these two genes in the caudal diencephalon and rostral mesencephalon of E11.5 mouse embryos (see below), this tissue was microdissected and used as a source of mRNA for Northern analysis and for the isolation of cDNAs. An 8.0-kb transcript for *Sim1* and a 4.2 kb transcript for *Sim2* were detected in this tissue by Northern analysis (data not shown). A cDNA library made from this tissue was screened for *Sim1* and *Sim2* cDNAs. A 4.0-kb clone containing the entire coding region of *Sim2* was isolated. Several overlapping cDNA clones and a genomic clone were used to assemble a 7.2-kb stretch of sequence of the *Sim1* cDNA containing the entire coding region of the gene (see Experimental Methods).

The bHLH and PAS Domains of the Drosophila sim Gene Product Are Highly Conserved in the Mouse

Figure 1 compares the deduced amino acid sequences of the two murine gene products to that of *Drosophila* SIM.

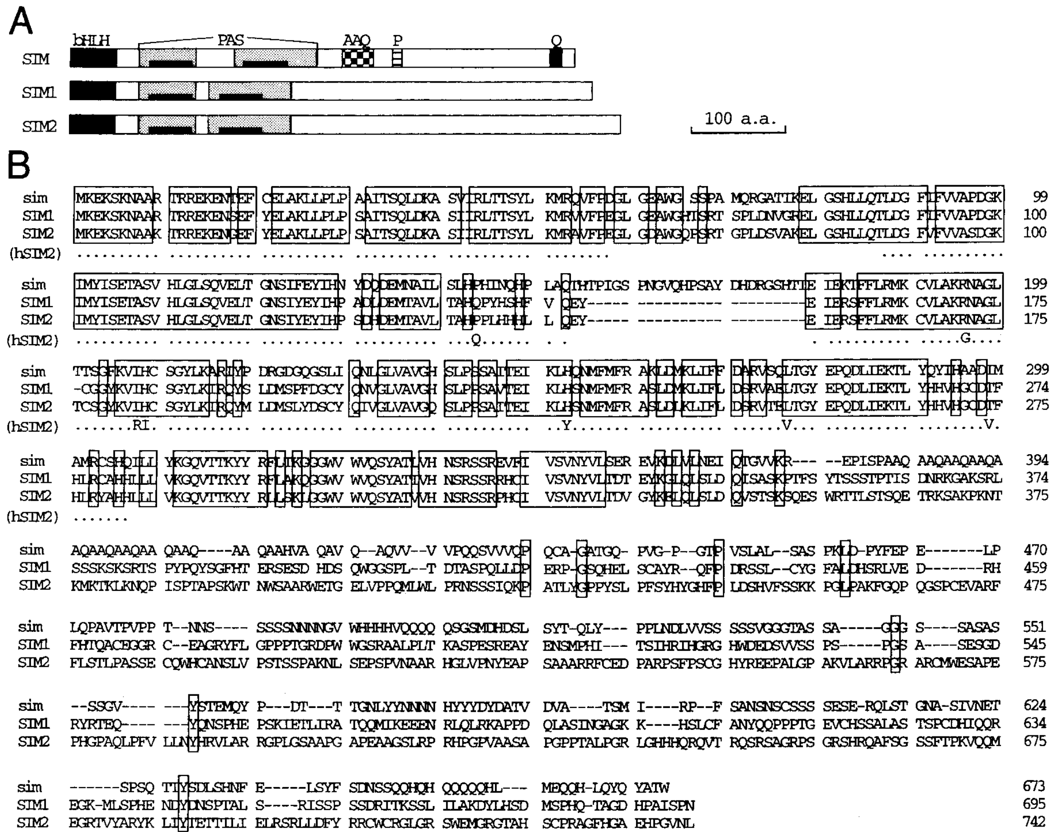


FIG. 1. Relationships between the predicted products of the *Drosophila sim* and the murine *Sim1* and *Sim2* genes. (A) Structural organization of the predicted SIM, SIM1, and SIM2 proteins. Each protein has a basic helix–loop–helix (bHLH) domain at its extreme N-terminus (dark shading). Each protein also has a PAS domain; the portions of the PAS domains that are conserved across species are indicated by boxes with light shading (the direct repeats within the PAS domain are indicated by black boxes). The three proteins diverge significantly after the PAS domain. Other structural motifs in the SIM protein are also illustrated: AAQ repeats (checker board), proline (P) rich region (horizonal lines), and a glutamine (Q) rich region (black box). (B) Alignment of the predicted amino acid sequences of the SIM, SIM1, and SIM2 proteins. Amino acids conserved between all three proteins are boxed. The nucleotide sequences of *Sim1* and *Sim2* have been submitted to GenBank (see Experimental Methods). The partial human SIM2 sequence (Chen *et al.*, 1995; Dahmane *et al.*, 1995) is aligned underneath the mouse SIM2 sequence: the differences between the predicted human SIM2 and mouse SIM2 amino acid sequences are shown, whereas dots indicate amino acids that are identical between human and mouse SIM2. Note the presence of two gaps in the human sequence compared to the murine seuqences.

The amino-terminal halves of the three proteins have similar structural domains (Fig. 1A). Like SIM, SIM1 and SIM2 have predicted bHLH domains at their extreme amino termini. The amino acid sequences of the bHLH domains are highly conserved between fly and mouse: there are only five substitutions in the SIM1 bHLH domain, and six in the SIM2 bHLH domain; for both genes, two of the substitutions are conservative (Fig. 1B).

Each of the murine proteins also has a PAS domain, which is highly homologous to that of *Drosophila* SIM. One difference between the murine and fly proteins is the presence, in the PAS domain of SIM, of a 24 amino acid insert (residues 155–178) between the two so-called PAS repeats (Fig. 1B). Interestingly, a hydrophobic region just amino terminal to the first PAS repeat that is conserved between PER, ARNT, AHR, and SIM (residues 89–94 of SIM: FIFVVA) is strongly conserved in SIM1 and SIM2. In particular, valine 92 of SIM is conserved in both murine proteins; mutation of the homologous valine in the PER protein to an aspartic acid in the *per*^{L} allele alters the biological function of PER and dramatically reduces the ability of the PER PAS domain to dimerize (Huang *et al.*, 1993). A second region that is very conserved between all PAS domain proteins (residues 277–283 of SIM: GYEPQDL) is identical in the fly and murine proteins.

The homology between the two murine proteins and SIM is also seen in the 40 amino-acid stretch linking the bHLH and PAS domains in each protein. Overall, throughout the amino terminal halves of the proteins (the first 375, 350, and 351 residues of SIM, SIM1, and SIM2, respectively), SIM1 and SIM2 are slightly more closely related to each other than they are to SIM at

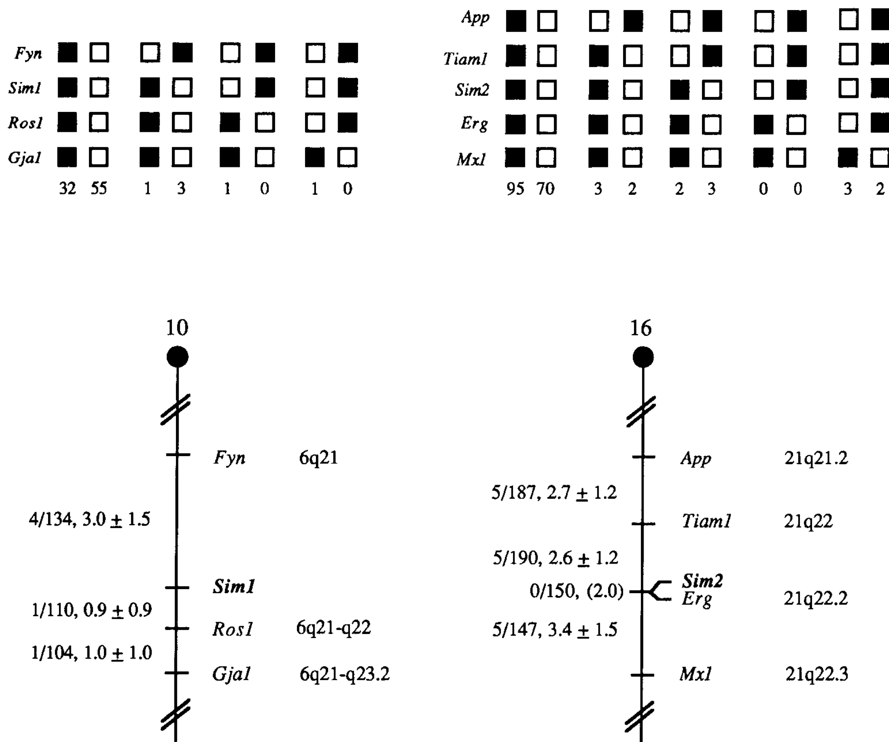
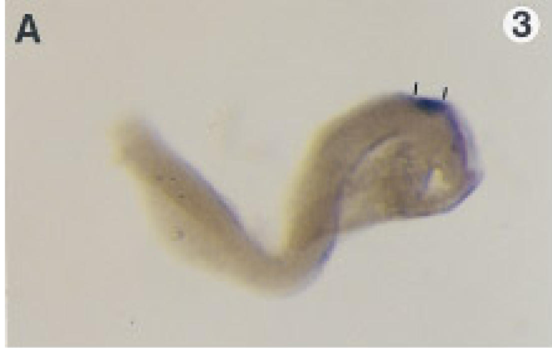


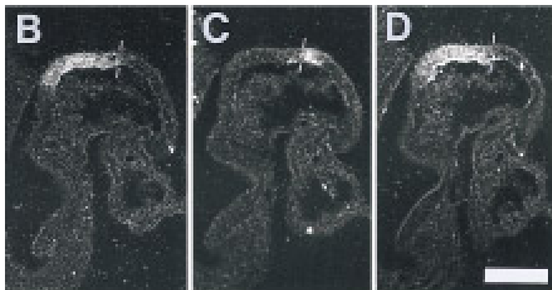
FIG. 2. Murine chromosomal location of *Sim1* and *Sim2.* The two genes were mapped to mouse chromosomes by interspecific backcross analysis. The segregation patterns of *Sim1, Sim2,* and the flanking genes in, respectively, 93 and 180 backcross animals that were typed for all loci are shown at the top of the figure. Each column represents the chromosome identified in the backcross progeny that was inherited from the (C57BL/6J × *M. spretus*)F1 parent. The black boxes represent the presence of the C57BL/6J allele and the white boxes represent the presence of the *M. spretus* allele. The number of offspring inheriting each type of allele is listed at the bottom of each column. Partial chromosome linkage maps showing the locations of the Sim loci in relation to linked markers (*Erg:* Rao *et al.*, 1987; *Fyn, Ros1,* and *Gja1:* Justice *et al.*, 1990; Haeflinger *et al.*, 1992; *App, Taim1,* and *Mx1:* Bae *et al.*, 1994; Habets *et al.*, 1995) are shown. The number of recombinant N2 animals over the total number of N2 animals typed plus the recombination frequencies expressed as genetic distance in centimorgans (±1 standard error) is shown for each pair of loci to the left of each chromosome. The upper 95% confidence limit of the recombination distance is given in parentheses when no recombinants are found between loci. Gene order was determined by minimizing the number of recombination events required to explain the allele distribution pattern. The positions of loci mapped to human chromosomes are shown to the right of the chromosome maps. References for the map positions of most human loci can be obtained from GDB (Genome Data Base), a computerized database of human linkage information maintained by the William H. Welch Medical Library of The Johns Hopkins University (Baltimore, MD).

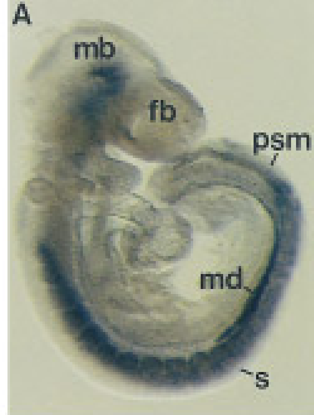
FIG. 3. Early expression pattern of *Sim2* in the forebrain primordium. (A) *Sim2* expression (bracketed by black lines) can be detected as early as the 2-somite stage (E8.0), as shown in a side view of an embryo whole-mount. (B–D) A comparison with *En1* expression shows that *Sim2* is expressed in a narrow band in the caudal portion of the forebrain, immediately rostral to the midbrain–forebrain boundary (indicated by the white lines). Three consecutive sections through a 10-somite (E8.5) embryo were hybridized with antisense probes to *En1* (B), *Sim2* (C), and *En1* again (D). The rostral limit of *En1* expression and the caudal limit of *Sim-2* expression are approximately at the same point (the white lines). At this stage, the rostral limit of *En1* expression defines the midbrain–forebrain boundary (McMahon *et al.*, 1992). Scale: 155 μm (A), 100 μm (B–D).

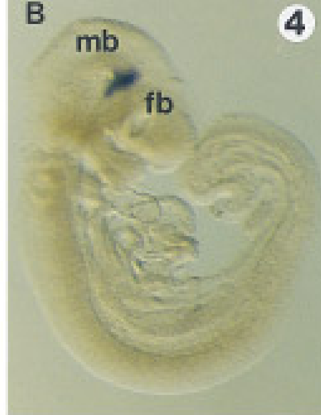
FIG. 4. Expression patterns of *Sim1* and *Sim2* in the E9.5 mouse. Transcripts were detected by *in situ* hybridization in embryo whole-mounts (A, B) and in tissue sections (C–F): *Sim1* expression is visualized in a lateral view of a whole-mounted embryo (A) and in a parasagittal section (C) and a transverse section (E) through the caudal diencephalon. At E9.5, *Sim1* expression is seen in the basal portion of the caudal diencephalon, the mesonephric ducts, and head mesenchyme. Note that expression in differentiated somites is restricted to the dermomyotome (see also Fig. 8). *Sim2* expression is visualized in a lateral view of a whole-mounted embryo (B) and a near midsagittal section (D) from the same embryo that was used in C. *Sim2* expression is confined to the caudal portion of the diencephalon in the basal plate and crosses the midline at the mammillary area (F). Weak expression in the mesonepheric tubules can also be detected by E9.5 (not shown). Note that in the mammillary area, *Sim1* is expressed in the basal half but is excluded from the midline (C), whereas *Sim2* is expressed in the midline (F). Abbreviations: dm, dermomyotome; fb, forebrain; hm, head mesenchyme; mb, midbrain; md, mesonephric duct; s, somite; psm, presomitic mesoderm. Scale, 420 (A, B), 360 (C, D), 100 (E, F) μm.

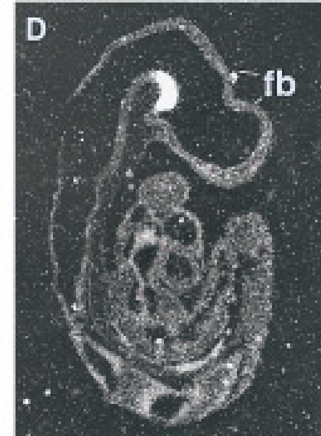
FIG. 5. Expression of *Sim1* compared to that of *Dlx2* at E10.5, as detected by radioactive *in situ* RNA hybridization of tissue sections. (Top) *Sim1* expression; (bottom) *Dlx2* expression. (A, B) Parasagittal sections. (C–N) Cross sections. Planes of cross sections in C–N are shown in A and B, as follows. Plane w: sections C–H; plane x: sections I and J; plane y: sections K and L; plane z: sections M and N. The boxed region in E is shown at fourfold higher magnification in F. The regions roughly corresponding to the p1/p2 and p2/p3 interprosomeric boundaries are indicated in F. G and H show details of the diencephalon from C and D, respectively. *Sim1* is expressed in two domains; domain S1 is in the ventricular zone and mantle of the basal plate of the diencephalon (BP DI) and mesencephalon (A, I, K, M), whereas domain S2 (a separate hypothalamic domain) is in the mantle (see arrows in A, K, M). Domain S1 has a dorsal extension at the zona limitans (zl; A, C, G). *Sim1* is also expressed in a restricted region of head mesenchyme (hm; A, M). *Dlx2* is expressed in two domains (see Bulfone *et al.*, 1993). Briefly, domain D1 is largely an alar plate zone that begins just anterior of the zona limitans (zl), extends along a longitudinal strip of hypothalamus (labeled as D1 in B, L, N; the domain extends from the ventral thalamus (VT), to the anterior midline just ventral to the optic stalk). Domain D2 is in the basal telencephalon and includes the medial and lateral ganglionic eminences (MGE, LGE) and preoptic area (POA) (B, D, H, J, L, N). There is a

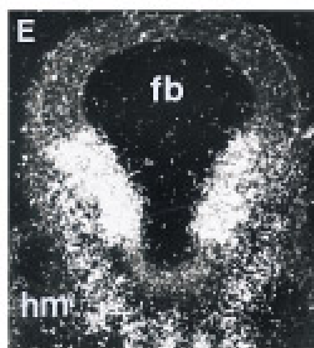


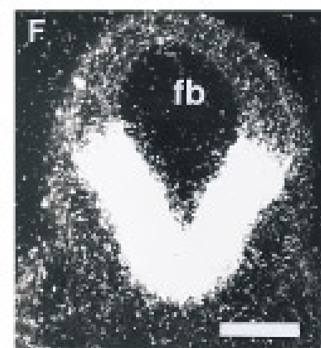


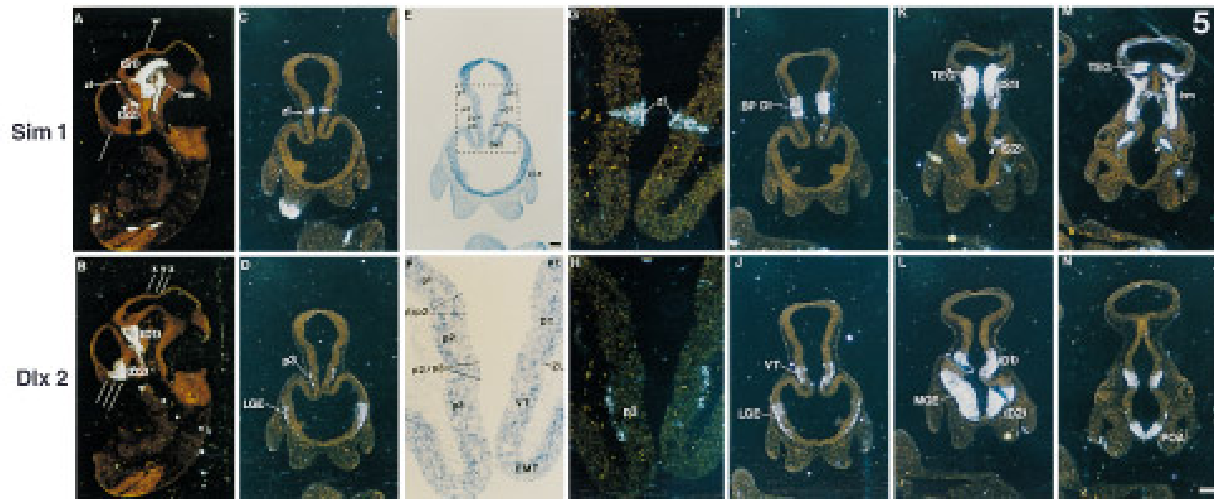












complicated pattern of expression in the ventricular and mantle zones (e.g., see D, J). Other abbreviations: DT, dorsal thalamus; EMT, eminentia thalami; PT, pretectum; TEG, tegmentum. Scale, 200 μm.

the amino acid level (SIM:SIM1, 69%; SIM:SIM2, 65%; SIM1:SIM2, 86%). However, no significant amino acid (19%) or nucleotide homology is seen among the carboxy-terminal halves of the three proteins (after their PAS domains). In particular, neither SIM1 nor SIM2 appears to have a glutamine-rich region like those found in SIM, ARNT, and AHR (Fig. 1A), which appear to act as transcriptional activation domains (Reisz-Porszasz *et al.*, 1994; Jain *et al.*, 1994; Franks and Crews, 1994).

Chromosomal Locations of the Sim1 and Sim2 Genes

The chromosomal locations of the murine *Sim1* and *Sim2* genes were determined using an interspecific backcrossing panel derived from crosses of [(C57BL/6J × Mus Spretus)F1 × C57BL/6J)] mice (Copeland and Jenkins, 1991; and unpublished; see Experimental Methods). Informative restriction-fragment-length polymorphisms (RFLPs) were used to map these two genes. *Sim1* mapped to the proximal region of mouse chromosome 10, 3 cM distal of *Fyn,* and 0.9 and 1.9 cM proximal of *Ros* and *Gjai1,* respectively (Fig. 2). Several mutations map in this interval, including *jackson circler* (*jc*), *waltzer* (*v*), *downless* (*dl*), and *kidney disease* (*kd*) (Green, 1989). *Sim2* mapped to the very distal end of chomosome 16. *App* and *Tiam1* map 5.3 and 2.6 cM proximal, respectively. *Sim2* is 3.4 cM proximal of *Mx1,* and has an identical strain distribution as *Erg.* The only mutation that maps in this vicinity is *weaver* (*wv*) (Fig. 2). According to the mouse–human linkage homologies, it is likely that *Sim1* maps to human 6q21 and *Sim2* to 21q22. Indeed, using this murine *Sim2* cDNA as a probe, a human *SIM2* gene has been mapped to this region of chromosome 21 (21q22.2–q22.3) (Muenke *et al.*, 1995), which corresponds to the Down syndrome critical region. Similarly, exon-trapping studies in this region have recently led to the isolation of six exons of a human *sim*-related gene (Chen *et al.*, 1995; Dahmane *et al.*, 1995). The ORF of these exons codes for 252 amino acids of a SIM homolog; the predicted sequence is highly homologous to the amino terminal 40% of the murine SIM2 sequence (Fig. 1B), with identical bHLH domains and only seven amino acid substitutions in the remainder of the human sequence. There are, however, two short gaps in the human sequence compared to the murine SIM2 sequence (Fig. 1B). These may represent bona fide differences; alternatively, it is possible that one or more intervening exons in this region of the human gene have not been identified.

Expression of Murine Sim Genes During Embryonic Development

The expression of *Sim1* and *Sim2* was studied by RNA *in situ* hybridization. Because *sim* plays a particularly important role at early stages of neurogenesis in *Drosophila,* we first describe the expression patterns of the two murine genes in the developing nervous system. Expression outside the nervous system is discussed later.

The Murine Sim Genes Are Expressed Early with Restricted Patterns in the Developing Forebrain

Sim2 is a marker of early regionalization of the anterior neural plate. The earliest expression of *Sim2* was observed in the anterior neural plate in 2-somite stage embryos (E8.0; Fig. 3A); it was not detected at E7.5 (data not shown). *Sim2* expression is restricted to a narrow transverse band of cells located near the junction of the midbrain (mesencephalon) and forebrain (prosencephalon). This localization was confirmed by comparing the expression of *Sim2* and *En1.* In serial sections of a 10-somite embryo, the caudal boundary of *Sim2* expression was found roughly to coincide with the rostral border of *En1* expression (Figs. 3B–3D), which at this stage is known to approximate the mesencephalon–prosencephalon boundary (McMahon *et al.*, 1992). This zone of expression is also restricted in the mediolateral dimension, being excluded from the lateral edge of the neural plate (data not shown). By E9.5 (24-somite stage), *Sim2* expression remains restricted to a ventral domain of the caudal diencephalon (Figs. 4B and 4D), that extends anteriorly into the mammillary region, where *Sim2* expression crosses the ventral midline (Fig. 4F).

Sim1 is also expressed in the caudal diencephalon. *Sim1* expression overlaps partially with that of *Sim2* in the caudal diencephalon, but has a slightly later onset (E9.0; Figs. 4A, 4C, and 4E show expression at E9.5). The rostrocaudal extent of expression is similar to that of *Sim2,* although it appears to extend slightly more caudally into the mesencephalon (Figs. 4A and 4C). As seen in transverse sections across the diencephalon, *Sim1* expression is also restricted to the basal half of the caudal diencephalon, but, unlike *Sim2,* is excluded from the ventral midline (Fig. 4E). In addition to expression in neural tube, *Sim1* is also expressed in the mesenchyme underlying the neural tube in the cephalic flexure (Figs. 4C, 4E).

The Sim Genes and Dlx2 Are Expressed in Adjacent Domains during Regionalization of the Forebrain

It has been proposed that the forebrain is divided into several neuromeres (Bulfone *et al.*, 1993; Figdor and Stern, 1993) called prosomeres (Puelles and Rubenstein, 1993; Rubenstein *et al.*, 1994). According to this model, the forebrain is subdivided by transverse (neuromeric) and longitudinal boundaries that separate distinct histo-

genic domains. A useful molecular marker for some of these boundaries is the *Dlx2* homeobox gene (Bulfone *et al.*, 1993). To define the boundaries of *Sim* gene expression in the forebrain when the initial regionalization of the forebrain is occurring, we compared *Sim1*, *Sim2* and *Dlx2* expression in serial sections of E10.5 and E12.5 embryos. Figure 5 shows some examples from this analysis, the results of which are summarized in schematic form in Fig. 6. The major features of the expression are described in the text; additional details are in the legends to these figures. As *Sim1* and *Sim2* were found to have similar expression patterns at E10.5 and E12.5, Fig. 5 shows expression data for only one of these genes (*Sim1*).

By E10.5, *Sim1* expression in the brain is found in two separate domains. The caudal domain (domain S1 in Fig. 5, which derives from the region where expression is observed at E9.5) is largely a longitudinal basal plate zone that extends rostrally from the midbrain/isthmic boundary, through the midbrain, prosomere 1 (p1, synencephalon), p2 (posterior parencephalon), p3 (anterior parencephalon), and the mammillary region (MA, basal p4) (Figs. 5 and 6). Domain S1 has a thin transverse extension at the p2/p3 boundary (zona limitans) (Figs. 5A, 5C, 5G, and 6). In addition, *Sim1* is expressed in the mantle of the isthmus (Fig. 6). The second *Sim1* brain expression domain (domain S2) is in the alar plate of the secondary prosencephalon. This domain is just forming at E10.5 (see arrows in Figs. 5A and 5K) and is fully formed by E12.5 (Fig. 6). Whereas at E10.5 *Sim1* expression in domain S1 largely includes both the ventricular (proliferative) and mantle (postmitotic) zones, its expression in domain S2 is detected only in the mantle zone.

The transverse extension of domain S1 at the p2/p3 boundary (zona limitans) is interesting because it coincides with a morphological landmark. It is well known that there are structural constrictions in the wall of prosencephalon at E10.5 that approximate the position of the boundaries between the prosomeres, including the p2/p3 boundary (Figs. 5E and 5F). Each constriction comprises an intraventricular ridge and a pial furrow, structures also found at interrhombomeric boundaries in the hindbrain (Heyman *et al.*, 1993). As shown in Figs. 5E–5G, *Sim1* is expressed precisely in a V-shaped group of cells that spans the region between the ridge and the furrow at the p2/p3 boundary, thus defining this boundary.

The expression pattern of *Sim2* is largely identical to that of *Sim1* (including at the p2/p3 boundary) but with some differences (Fig. 6), the most salient of which are as follows: (i) *Sim1* expression extends caudal to that of *Sim2* in both domain S1 and domain S2; (ii) only *Sim2* expression crosses the ventral midline in the mammillary area (as discussed above at E9.5); (iii) at E12.5, only *Sim1* is expressed in the telencephalon (see Fig. 6 legend for details).

Dlx2 is also expressed in two domains (domains D1 and D2 in Fig. 5; Porteus *et al.*, 1991; Robinson *et al.*, 1991; Bulfone *et al.*, 1993; Porteus *et al.*, 1994), which are nonoverlapping with the two *Sim* expression domains, but which abut these expression domains at several boundaries in both the transverse and longitudinal planes (Figs. 5 and 6). In particular, the *Sim* and *Dlx2* expression domains are adjacent at the p2/p3 transverse boundary (Figs. 5A–5D, 5G, 5H, and 6), with *Sim1* and *Sim2* expressed in the cells of the p2/p3 boundary (see above), whereas *Dlx2* is expressed just anterior to the boundary in the ventral thalamus (p3) (Figs. 5B, 5D, 5H, and 6).

Later Sites of Expression in Brain

The expression patterns of the two *Sim* genes were also examined at E16.5. Expression was observed in brain regions that derived from the sites of earlier expression (Table 1; some expression data are visible in Fig. 9). Of particular interest, however, was the additional finding of expression of the *Sim2* gene in the cerebral cortex and olfactory bulb, though at low levels (Table 1 and data not shown).

Other Sites of Sim Gene Expression

Sim1 is expressed in the ventral midline region of the spinal cord after E10.0. Neither *Sim1* nor *Sim2* is expressed in the spinal cord before E9.5. By E10.0, however, *Sim1* is expressed in the spinal cord in bilateral stripes flanking the floor plate (Figs. 6, 7A, and 8G). The domain of *Sim1* expression appears to be immediately adjacent to the floor plate, as demonstrated by comparing the expression of *Sim1* to that of the floor plate marker *F-spondin* (Klar *et al.*, 1992) in adjacent sections (compare Figs. 7A and 7B). The site of *Sim1* expression corresponds to a poorly defined region which has been dubbed "area X" in the chick embryo (Yamada *et al.*, 1991); the cell types that arise from area X are not known, though they might include so-called "primitive longitudinal" neurons in the chick (Yaginuma *et al.*, 1990). This expression pattern in the ventral spinal cord is maintained through E16.5 (the latest time examined; data not shown).

Sim1 provides a marker for the regionalization of somites. High levels of *Sim1* expression are also observed in the paraxial mesoderm (Figs. 4A and 4C). *Sim1* is expressed initially uniformly in presomitic mesoderm (Fig. 8A). When somites first pinch off from the presomitic mesoderm, *Sim1* is still expressed in a uniform fashion. However, as each somite becomes patterned into dermo-

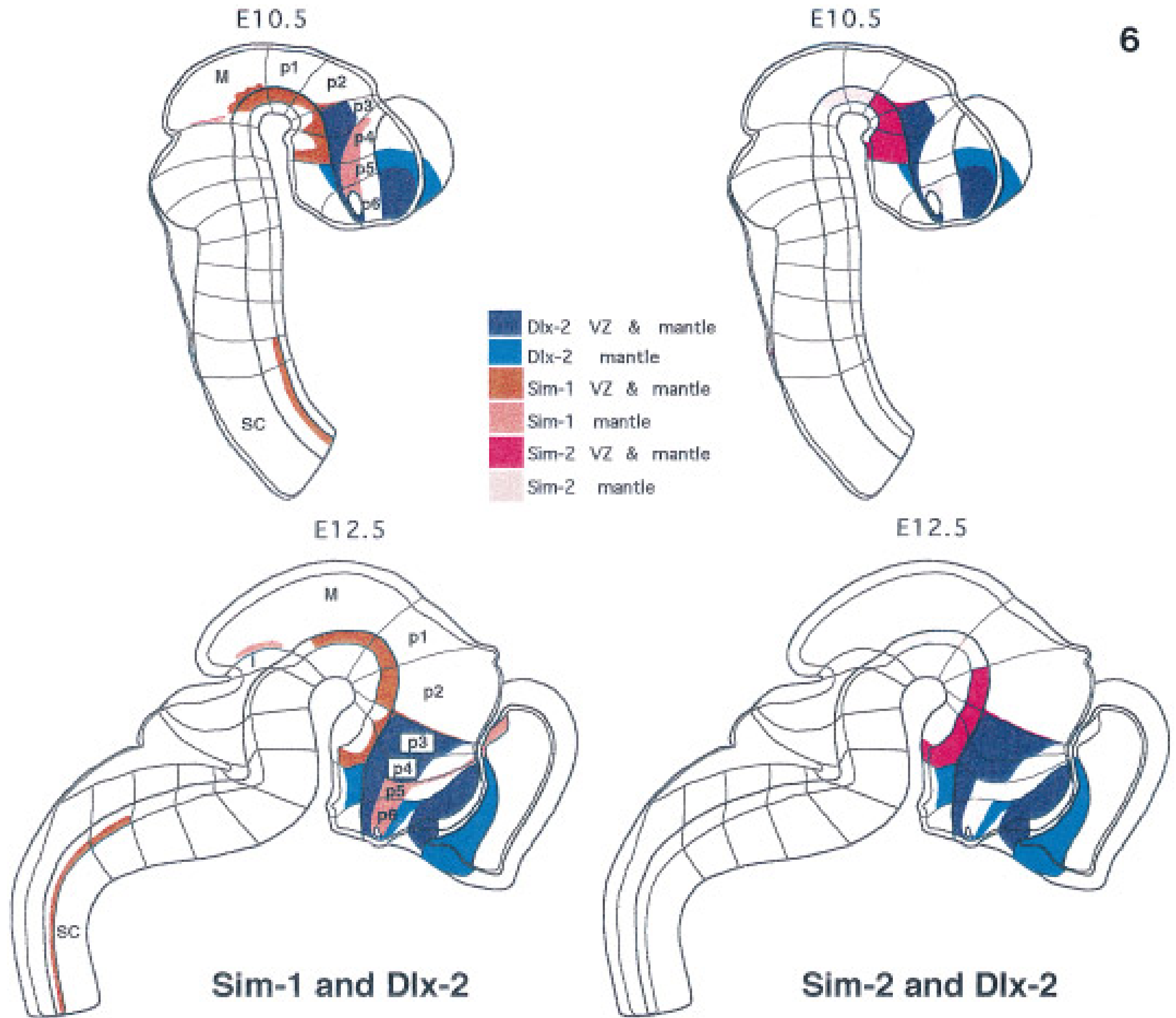


FIG. 6. Schema showing expression of *Sim1*, *Sim2*, and *Dlx2* at E10.5 and E12.5. Adjacent sections were used for this study to compare the expression of each gene in order to refine the map of their expression patterns. The domains of expression respect transverse and longitudinal boundaries that are incorportated within the framework of the Prosomeric model (Puelles and Rubenstein, 1993; Rubenstein *et al.*, 1994). Expressions in the mantle and ventricular zone (VZ) are indicated in different colors: *Sim1* is shown in dark orange (VZ) and pale orange (mantle), *Sim2* in magenta (VZ) and pink (mantle), and *Dlx2* in purple (VZ) and blue (mantle). Abbreviations: M, midbrain; p, prosomere; SC, spinal cord.

myotome and sclerotome, and subsequently into dermatome, myotome, and sclerotome, *Sim1* expression becomes restricted first to the dermomyotome and then the dermatome (Figs. 8C and 8E). Interestingly, this pattern of *Sim1* expression is the mirror image of that of a different bHLH protein-encoding gene, *Mtwist*, which is also expressed uniformly in presomitic mesoderm but becomes gradually enriched in the sclerotome (though it is not altogether excluded from the other two components of the somite, Figs. 8B, 8D, and 8F) (Wolf *et al.*, 1991). At later stages, *Sim1* expression becomes restricted to the lateral aspect of the dermatome (data not shown).

Expression in derivatives of somites. Between E9.5 and E10.5, *Sim1* is also expressed in cells migrating away from the ventrolateral margin of the dermatome, which, at the limb level, contain the presumed limb myoblast precursors (Chevallier *et al.*, 1977; Christ *et al.*, 1977). It is, however, detected at only very low levels in the limb proper (Fig. 8G). In contrast, from E10.5, *Sim2* is expressed strongly in the limbs in regions that appear to

TABLE 1
Expression Pattern of the *Sim* Genes at E16.5

Sim1	Sim2
Forebrain	Forebrain
Domain 1	Domain 1
Mammillary nuclei (lateral mammillary)	Mammillary nuclei (lateral mammillary)
Zona limitans	Zona limitans
Basal plate domains of p1, p2	Basal plate domains of p1, p2
Domain 2	Domain 2
Anterior hypothalamic nuclei	Anterior hypothalamic nuclei
Amygdala	Cortex
Midbrain	Olfactory bulb
Basal plate	Midbrain
Spinal cord	Basal plate
Cells next to the floor plate	Cartilage/bone
Head mesenchyme	Ribs
Dermis (face, chest, and back)	Vertebrae
Genital eminence	Palatal
Kidney tubules	Mandibular
	Hyoid
	Limb/digits
	Skeletal muscles
	Subsets of the trunk muscles at the limb level
	Oral epithelium
	Kidney tubules

correspond approximately to the dorsal and ventral muscles masses (Fig. 8H). At later stages (E12.5 and E16.5), *Sim1* is also found in a layer of cells immediately beneath the epidermis in the dorsal half of the embryo (data not shown); these cells are presumed to be dermal cells, which would be consistent with the earlier expression of *Sim1* in dermatome. At E10.5, high levels of *Sim2* expression in the mesenchymal region of all the branchial arches were also detected (data not shown). This is consistent with its later expression in the tongue, mandibular bones, and trachea (Table 1). *Sim2* is also expressed in the vertebrae, ribs, and a subset of muscles between E12.5 and E16.5 (the latest time point examined).

Expression in derivatives of the mesonephros and metanephros. *Sim1* is also expressed at high levels in the mesonephric ducts during their formation (Figs. 4A, 4B; 8A, 8C, 8E). Interestingly, *Sim2* is expressed in complementary fashion in the mesonephric tubules, although at lower levels (data not shown). Expression of *Sim1* and *Sim2* in the mesonephros persists at E10.5 and E11.5. At E12.5 and 16.5, *Sim1* is expressed in the cortical glomeruli (data not shown) and *Sim2* in the central tubules of the kidney (Figs. 9A, 9B).

Other sites of expression. These are listed in Table 1 and discussed below.

DISCUSSION

The Mouse Genome Contains Two Sim Homologs

In *Drosophila*, *sim* is a key regulator of the development of cells at the midline of the central nervous system. We have shown that the mouse genome contains at least two genes that are highly related to *sim*. The extensive homology between the three genes suggests that they are indeed evolutionary homologs. The high homology within the bHLH domains of all three proteins strongly suggests that they may have similar DNA binding and dimerization properties.

The three SIM proteins also share a 250-amino-acid-long PAS domain. This domain is conserved among the *Drosophila* SIM and PER proteins and the mammalian ARNT, AHR, and HIF-1 proteins (Nambu *et al.*, 1991, Wang *et al.*, 1995). The PAS domain of PER can homodimerize and can also form a heterodimer with the PAS domain of SIM (Huang *et al.*, 1993). Since PER does not contain a discernible bHLH domain, it has been proposed that PER binds other PAS domain-containing proteins and thereby interferes with their DNA binding function, thus acting as a negative regulator of these proteins. Consistent with this possibility, the SIM PAS domain when expressed alone has been shown to antagonize endogenous *sim* function *in vivo* (Franks and Crews, 1994). Residues within the PAS domain that are important for the dimerization function of PER (Huang *et al.*, 1993) are conserved in the mammalian SIM proteins, suggesting that the PAS domains in the mammalian proteins may indeed function in dimerization. In fact, since both HLH and PAS domains are thought to mediate dimerization, it has been proposed that the PAS domain may serve to consolidate dimerizations initiated by HLH

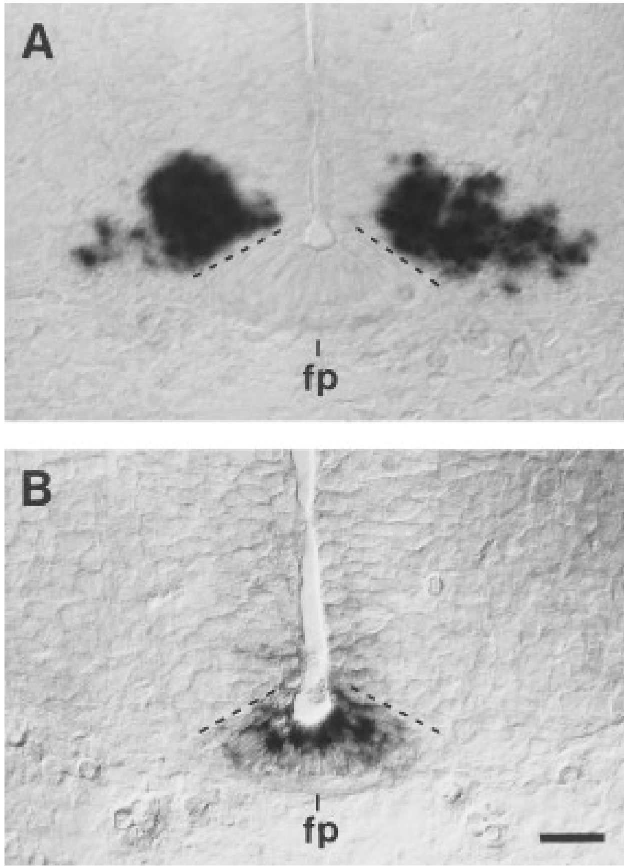


FIG. 7. *Sim1* is expressed in the ventral midline region of the spinal cord. (A) A high-power view of spinal cord in a transverse section through the trunk region of an E10.5 embryo. *Sim1* transcripts, detected using a digoxygenin-labeled probe, were observed in two symmetrical groups of cells that flank the floor plate of the spinal cord (A). These cells appear directly to abut the floor plate, which has been visualized in B by expression of the *F-spondin* gene, also detected nonisotopically in a nearby section of the same embryo. fp, floor plate. Scale, 10 μm.

domains (Huang *et al.*, 1993). Two other known functions of PAS domains of AHR are the binding of aromatic hydrocarbon ligands and binding to heat-shock protein 90 kDa (HSP90) (Wilhelmsson *et al.*, 1990; Burbach *et al.*, 1992; Whitelaw *et al.*, 1993). It is therefore tempting to speculate that the PAS domains of SIM proteins also mediate binding to specific ligands and association with heat shock proteins.

General Features of Mammalian Sim Expression Patterns

A general feature of the expression of the mammalian *Sim* genes is that some of their sites of expression roughly parallel those of *Drosophila sim*. In *Drosophila*, *sim* is expressed in the brain, ventral nerve cord, gut, and muscles (Crews *et al.*, 1988; Lewis and Crews, 1994). Similarly in mouse, *Sim1* is expressed in the brain, ventral spinal cord, and foregut, and *Sim2* is expressed in the brain and in muscles. In addition, both mammalian *Sim* genes are expressed in the developing mesonephros and, later, the kidney. Importantly, whereas *sim* is a key regulator of CNS midline cell development in the fly, the mammalian *Sim* genes are not expressed in floor plate cells of the spinal cord. The only expression of the mouse *Sim* genes in the ventral midline of the central nervous system that we have detected is *Sim2* expression in the mammillary area of the diencephalon.

Sim2 Is an Early Marker of Regionalization of the Embryonic Forebrain

Sim2 is expressed in the forebrain in a restricted pattern in the caudal diencephalon at early stages of neural development (by the two-somite stage). The importance of this expression pattern is in showing that the forebrain is already regionalized at the two-somite stage, prior to any overt morphological specialization of the diencephalon primordium. This observation is consistent with the recent results of Shimamura *et al.* (1995), defining the earliest stages of forebrain regionalization.

Sim1 Is a Marker of Regionalization of Somites

Although *Sim2* is expressed in many tissues derived from somites by E12.5 (including muscle and cartilage), it is not detected in somites. In contrast, *Sim1* is detected from the earliest stages of somite development. *Sim1* is first expressed uniformly in the presomitic mesoderm and becomes restricted to the dorsolateral compartment of the somite, the dermamyotome, as each somite matures. During the elaboration of the dermatome and myotome, *Sim1* expression is restricted to the dermatome. Interestingly, *Sim1* is also detected in cells that migrate away from the ventrolateral lip of the dermomyotome, which are known to include cells that migrate into limbs where they become myoblasts. *Mtwist*, which also encodes a bHLH-containing protein, is also expressed uniformly in the presomitic mesoderm, but is then upregulated in the sclerotome and down-regulated in the dermamyotome during the maturation of the somite (Wolf *et al.*, 1991). *Myf5*, also a gene encoding a bHLH protein, is expressed in the presomitic mesoderm, albeit at lower levels. *Myf5* is subsequently detected in the dorsal–medial lip of the somite where the myotome precursor cells reside, and later in the mature myotome (Ott *et al.*, 1991). Thus, the regionalization of the undifferentiated somite into dermatome, myotome, and sclerotome is mirrored

by the restriction of these three genes encoding bHLH proteins to the three regions.

Sim2 and Down Syndrome

Down syndrome, the most common human birth defect, arises from trisomy of chromosome 21. A region of chromosome 21 that is critical for this syndrome has been defined by analysis of cases of partial trisomy 21, leading to the definition of the critical region around 21q22.2, duplication of which causes many of the phenotypes of the syndrome (Delabar *et al.*, 1993). Based on chromosomal location and sequence homology, the human *sim* homolog identified within the critical region corresponds to the mouse *SIM2* gene (Chen *et al.*, 1995; Muenke *et al.*, 1995; Dahmane *et al.*, 1995).

A recent study on the expression pattern of the *SIM2* gene in human and rat embryos using one human *SIM2* exon as a probe (Dahmane *et al.*, 1995) is consistent with the data presented here. The early expression of the murine *Sim2* gene in the forebrain, craniofacial structures, the axial skeleton, and a subset of muscle cells (Figs. 3, 4, 6, and 9 and Table 1), is certainly consistent with a possible involvement of the human *SIM2* gene in causing the phenotypic characteristics of Down syndrome patients, which include mental retardation, craniofacial defects, short stature, and hypotonia. If *SIM2* does indeed play a role, this would presumably be a consequence of increased *SIM2* expression associated with full or partial trisomy 21. Future studies involving manipulations of *Sim2* gene expression in mice should help address this possibility.

EXPERIMENTAL METHODS

Isolation of Mouse Genomic Clones

One million recombinants of an NIH3T3 mouse λ-EMBL genomic library (gift of Dr. R. Grosschedl) were screened with a ^{32}P-labeled probe for the sequences encoding the bHLH domain of the *Drosophila sim* gene. The probe was generated by PCR as described (Maisonpierre *et al.*, 1989) using a *sim* cDNA as a template (gift of Dr. S. Crews) and sequence specific primers based on the *sim* sequence (Nambu *et al.*, 1991). Filters (Hybond-N, Amersham) were hybridized at 60°C in 5× SSC, 5× Denhardt's, 0.5% SDS, 250 μg/ml salmon sperm DNA, and 2×10^6 cpm/ml of the probe for 24 h. The filters were washed repeatedly at 55°C in 2× SSC, 1% SDS and autoradiographed. Seven hybridizing clones were isolated and digested with a variety of restriction enzymes. For each clone, a hybridizing fragment was identified by Southern analysis, subcloned into pBluescript, and sequenced. Each clone contained a 171-nt sequence highly homologous to the sequence encoding the bHLH domain of *sim.* Five clones contained fragments of one gene, designated *Sim1* gene; the other two contained fragments of a second *sim*-related gene, designated *Sim2.* In both genes, the 171-nt stretch homologous to *sim* is located at the extreme 5′ end of the coding region (see below) and, as in the *sim* gene, is flanked 3′ by an intron.

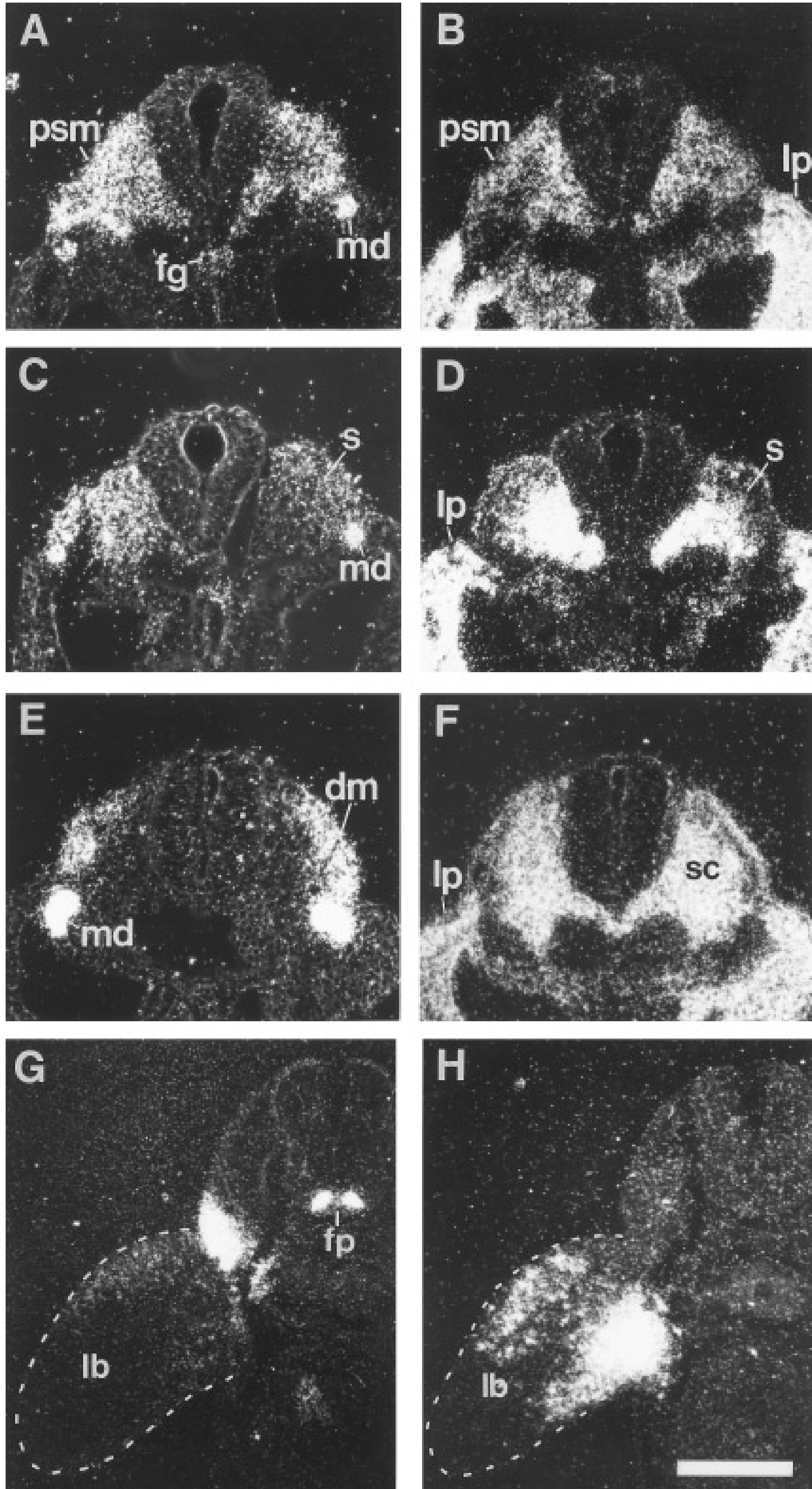
Northern Blot Analysis

After RNA *in situ* hybridization studies showed expression of both *Sim1* and *Sim2* in the diencephalon and mesensephalon at E11.5, portions of the E11.5 head containing diencephalon and mesencephalon were microdissected, and total RNA was extracted (Chomczynski and Sacchi, 1987). Poly A^+ RNA was isolated by oligo-(dT) cellulose (Sambrook *et al.*, 1989). For Northern analysis, 10 μg of polyA^+ RNA was used (Sambrook *et al.*, 1989). The blotted membrane was hybridized with a ^{32}P-labeled genomic fragment of *Sim1* and *Sim2,* washed, and autoradiographed for 10 and 3 days, respectively, at −80°C in the presence of an intensifying screen. These probes detected an 8.0-kb *Sim1* transcript and a 4.2-kb *Sim2* transcript.

Isolation of cDNA Clones

A λ-uni-ZAP-XR cDNA library (Stratagene) was made according to manufacturer's instructions using E11.5 mouse head polyA^+ RNA. Because the genomic probes for *Sim1* and *Sim2* only contained exons for very 5′ sequences in the *Sim1* and *Sim2* cDNAs (see above), we first used 3′ RACE (Frohman *et al.*, 1990) to isolate more 3′ sequences. Using E11.5 mouse head cDNA as a template, we isolated a 1.2-kb *Sim1* cDNA and a 0.8-kb *Sim2* cDNA. These were used as probes to screen 4×10^6 recombinants from the mouse head library (Sambrook *et al.*, 1989). Eleven independent *Sim2* cDNAs were isolated. One of these was sequenced and found to contain the full-length *Sim2* coding region. The predicted initiating methionine in the inferred SIM2 protein is the first residue of the bHLH domain (Fig. 1).

By sequential walking using overlapping *Sim1* probes, the library screens also yielded five partial *Sim1* cDNAs covering about 7.2 kb of *Sim1* cDNA sequence. The assignment of the translation initiation site was verified by both the genomic sequences and cDNA sequences amplified by RT–PCR across an intron junction. As in SIM2 and *Drosophila* SIM, the inferred initiating methio-


A
psm
fg
md
B
psm
lp
C
s
md
D
lp
s
E
dm
md
F
lp
sc
G
fp
lb
H
lb

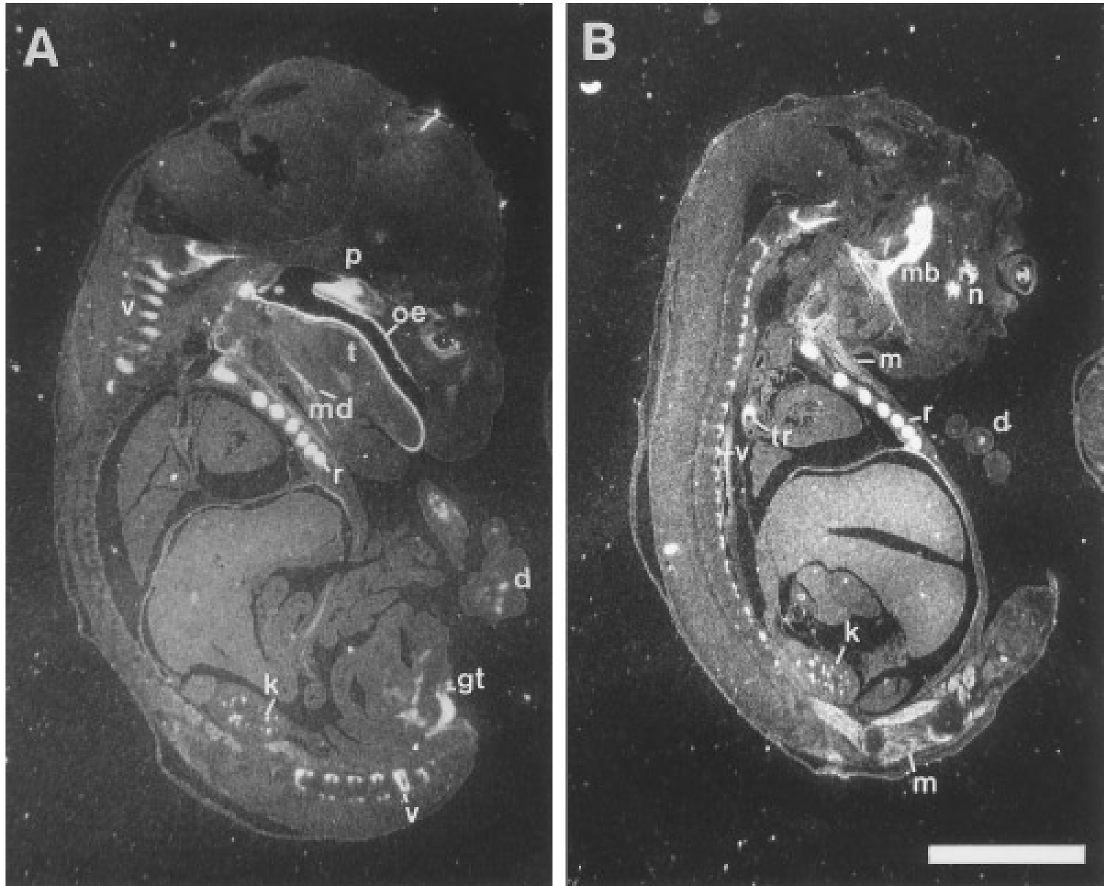


FIG. 9. Expression of *Sim2* detected in a midsagittal (A) and a parasagittal (B) section of an E16.5 mouse. *Sim2* is expressed in vertebrae (v) and ribs (r), the central tubule region of the kidney (k), the oral epithelium (oe), mandible (md), mandibular bone (mb), palate (p), tongue (t), nasal pit (n), trachea (tr), muscles (m), and the digits (d) of the limb. Scale, 2 μm.

nine in the SIM1 protein is the first residue of the bHLH domain (Fig. 1).

DNA Sequence Analysis

Fragments of *Sim1* and *Sim2* genomic and cDNA clones were subcloned into pBluescript KS- and SK- (Stratagene). *Exo*III nuclease deletions were performed on these clones in both directions; nested deletions were sequenced using T3 or T7 primers by the chain termination method (Sanger *et al.,* 1977) using Sequenase (USB). Nucleotide and protein sequence analysis was performed using Geneworks (Intelligenetics). Sequences have been deposited in the Genbank database with the

FIG. 8. *Sim1* and *Mtwist* are complementary markers of somite regionalization. The expression of *Sim1* (A–C) and *Mtwist* (D–F) during the regionalization of somites is shown in transverse sections taken from a single E9.5 mouse embryo, at the level of presomitic mesoderm (A, D), the most recently segmented somite (B, E), and a slightly older somite (C, F). (A, D) In the presomitic mesoderm, *Sim1* (A) and *Mtwist* (D) are expressed uniformly. (B, E) In the most recently segmented somite of this embryo, *Sim1* is gradually excluded from the ventromedial half of the somite (right-hand somite in B), presaging the division of the somite into dermomyotome and sclerotome. (Note that the section, being slightly oblique, runs through the left-hand somite at a more caudal level, where *Sim1* expression is not yet excluded from the ventromedial portion.) Conversely, *Mtwist* is enriched in the ventromedial portion of the somite (E). (Note that *Mtwist* expression is not completely extinguished in the dorsolateral half.) (C, F) At slightly more rostral levels where a differentiated dermomyotome is seen, *Sim1* is expressed in the dermomyotome but not the sclerotome, whereas *Mtwist* is highly enriched in the sclerotome (but is still expressed at lower levels in the dermomyotome). At all these levels, *Sim1* is also expressed in the mesonephric ducts and the dorsal aspect of the gut, and *Mtwist* is also expressed in the lateral plate mesoderm. (G, H) Expression of *Sim1* and *Sim2* at forelimb levels of an E10.5 embryo. (G) *Sim1* is expressed in the girdle, as well as in two bands of cells flanking the floor plate of the neural tube (see Fig. 7). Low levels of *Sim1* expression are observed in the dorsal aspect of the limb. (H) *Sim2* is expressed in dorsal and ventral aspects of the limb, apparently in the region of the muscle masses. dm, dermomyotome; fg, foregut; fp, floor plate; lp, lateral plate mesoderm; md, mesonephric duct; s, undifferentiated somite; psm, presomitic mesoderm; lb, limb. Scale, 100 μm (A–F), 134 μm (G, H).

following accession numbers: *Sim1:* U40575; *Sim2:* U40576.

Interspecific Mouse Backcross Mapping

Interspecific backcross progeny were generated by mating (C57BL/6J × *Mus Spretus*)F1 females and C57BL/6J males as described (Copeland and Jenkins, 1991). A total of 205 N2 mice were used to map the *Sim1* and *Sim2* loci. DNA isolation, restriction enzyme digestion, gel electrophoresis, Southern blot, and hybridization were performed as described (Jenkins *et al.*, 1982). ^{32}P-labeled probes were used. The *Sim1* probe, a 5.6-kb *Eco*RI/*Xho*I fragment of mouse cDNA, detects a 12.0-kb fragment in C57BL/6J DNA and fragments of 9.7 and 2.6 kb in *M. spretus* DNA following *Xba*I digestion. The *Sim2* probe, a 1.7-kb *Bam*HI/*Xho*I fragment of mouse cDNA, detects a 7.9-kb fragment in C57BL/6J and a 8.6-kb fragment in *M. spretus* DNA after *Bam*HI digestion. Neither probe contains the bHLH or the PAS domain. The presence or absence of the *M. spretus*-specific fragments were followed in backcross mice. The map locations of several other loci used to position *Sim1* and *Sim2* on our interspecific backcross have been previously described (see Fig. 2). Recombination distances were calculated as described (Green, 1981) using the computer program SPRETUS MADNESS. Gene order was determined by minimizing the number of double recombination events required to explain the allele distribution patterns.

In Situ Hybridization to RNA in Tissue Sections

Embryos from different stages were fixed in 4% paraformaldehyde in phosphate-buffered saline, embedded in paraffin, and sectioned at 8-μm thickness as described (Frohman *et al.*, 1990). RNA probes for *in situ* hybridization were transcribed from linearized plasmids using T7 RNA polymerase in the presence of [^{35}S]UTP. The *Sim1* probe was an antisense transcript spanning 171 nt of coding and 582 nt of 5′ untranslated and intron sequences; the *Sim2* probe was a 450-nt antisense probe spanning 171 nt of coding and 282 nt of 5′ untranslated and intron sequences. Antisense probes containing just the 5′ untranslated and intron sequences or just the sequences encoding the bHLH domains of each gene gave lower signals but identical hybridization patterns (data not shown). Sense probes transcribed using T3 RNA polymerase did not show any specific hybridization (data not shown). The *En1* cDNA probe (Joyner, 1988), the *Mtwist* cDNA probe (Wolf *et al.*, 1991), the *Fspondin* cDNA probe (Klar *et al.*, 1992), and the *Dlx* 2 probe (Bulfone *et al.*, 1993) were as previously described. For each probe, sections were processed, prehybridized, hybridized, and washed as described (Frohman *et al.*, 1990). Slides were coated with photographic emulsion (K-5, Polysciences, Inc.), developed after 3–4 weeks, stained with hematoxylin and eosin, dehydrated, cleared with Xylene, mounted in Permount (Fisher), dried, and examined by bright-field and dark-field illumination.

For hybridization to tissue sections with digoxygenin–UTP labeled probes, the method of Schaeren-Wiemers and Gerfin-Moser (1993) was used exactly as described with the exception that cryosections were taken from embryos fixed in 4% paraformaldehyde. Embryos were embedded in OCT embedding compound (Tissue-Tek) and sectioned at 12 μm. Antisense RNA probes were transcribed in the presence of digoxygenin–UTP (Boehringer Mannheim) using T7 polymerase as recommended by the manufacturer.

In Situ Hybridization to RNA in Embryo Whole-Mounts

The protocol for whole-mount *in situ* hybridization was based on the procedure of Wilkinson and Nieto (1993), as modified by Dr. J. McMahon (personal communication). The *Sim1* probe contained 500bp of the 5′ untranslated region and intron sequences and 1.2 kb of the coding sequence; the *Sim2* probe contained 300 bp of 5′ untranslated region and intron sequences and 1 kb of coding sequence. The color reaction (NBT and BCIP, GIBCO–BRL) was allowed to proceed for 5–24 h for *Sim2*, and 12–16 h for *Sim1*.

ACKNOWLEDGMENTS

We thank T. Jessell, M. Placzek, and G. Martin for helpful comments on an early version of this manuscript, D. J. Gilbert for excellent technical assistance, and Drs. A. Klar, A. Joyner, and F. Perrin-Schmidt for the *Fspondin*, *En1*, and *Mtwist* cDNAs, respectively. This work was supported by grants to M.T.L. from the Lucille P. Markey Charitable Trust and the Human Frontiers Science Program (Grant RG41/95), and by the Howard Hughes Medical Institute. C.-M. Fan was supported by postdoctoral fellowships from the Damon Runyon–Walter Winchell Cancer Research Fund, the American Cancer Society, California Division, and the Howard Hughes Medical Institute. This work was also supported by grants to J.L.R.R. (Human Frontiers Science Program Grant RG41/95, March of Dimes, NARSAD, the John Merck Fund, Pfizer Pharmaceuticals, and NIMH RO1 MH49428-01, NIMH RO1 MH51561-01A1, and K02 MH01046-01) and, in part, by the National Cancer Institute, DHHS, under contract with ABL. M.T.L. is an Assistant Investigator of the Howard Hughes Medical Institute.

REFERENCES

Bae, S.-C., Ogawa, E., Maruyama, M., Oka, H., Satake, M., Shigesada, K., Jenkins, N. A., Gilbert, D. J., Copeland, N. G., and Ito, Y. (1994).

PEBP2aB/mouse AML1 consists of multiple isoforms that possess differential transactivation potential. *Mol. Cell. Biol.* **14:** 3242–3252.

Bulfone, A. P., Puelles, L., Porteus, M. H., Frohman, M. A., Martin, G. R., and Rubenstein, J. L. (1993). Spatially restricted expression of Dlx-1, Dlx-2 (Tes-1), Gbx-2, and Wnt-3 in the embryonic day 12.5 mouse forebrain defines potential transverse and longitudinal segmental boundaries. *J. Neurosci.* **13:** 3155–3172.

Burbach, K. M., Poland, A., and Bradfield, C. A. (1992). Cloning of the Ah-receptor cDNA reveals a distinctive ligand-activated transcription factor. *Proc. Natl. Acad. Sci. USA* **89:** 8185–8189.

Chen, H., Chrast, R., Rossier, C., Gos, A., Antonarakis, S. E., Kudoh, J., Yamaki, A., Shindoh, N., Maeda, H., Minoshima, S., and Shimizu, N. (1995). Single-minded and Down syndrome. *Nature Genet.* **10:** 9–10.

Chevallier, A., Kieny, M., and Mauger, A. (1977). Limb–somite relationship: Origin of the limb musculature. *J. Embryol. Exp. Morph.* **41:** 245–258.

Chomczynski, P., and Sacchi, N. (1987). Single-step method of RNA isolation by acid gaunidinium thiocyanate-phenol-chloroform extraction. *Anal. Biochem.* **162:** 156–159.

Christ, B., Jacob, H. J., and Jacob, M. (1977). Experimental analysis of the origin of the wing musculature in avian embryos. *Anat. Embryol.* **150:** 171–186.

Copeland, N. G., and Jenkins, N. A. (1991). Development and application of a molecular genetic linkage map of the mouse genome. *Trends Genet.* **7:** 113–118.

Crews, S., Franks, R., Hu, S., Matthews, B., and Nambu, J. (1992). Drosophila single-minded gene and the molecular genetics of CNS midline development. *J. Exp. Zool.* **261:** 234–244.

Crews, S., Thomas, J., and Goodman, C. (1988). The *Drosophila single-minded* gene encodes a nuclear protein with sequence similarity to the *per* gene product. *Cell* **52:** 143–151.

Dahmane, N., Charron, G., Lopes, C., Yaspo, M.-L., Maunoury, C., Decorte, L., Sinet, P.-M., Bloch, B., and Delabar, J-M. (1995). Down syndrome critical region contains a gene homologous to Drosophila sim expressed during rat and human central nervous system dvelopment. *Proc. Natl. Acad. Sci. USA* **92:** 9191–9195.

Davis, C. A., and Joyner, A. L. (1988). Expression patterns of the homeo box-containing genes En-1 and En-2 and the proto-oncogene int-1 diverge during mouse development. *Genes Dev.* **2:** 1736–1744.

Delabar, J. M., Theophile, D., Rahmani, Z., Chettouch, Z., Blouin, J. L., Prieur, M., Noel, B., and Sinet, P. M. (1993). Molecular mapping of twenty-four features of Down syndrome on chromosome 21. *Eur. J. Human Genet.* **1:** 114–124.

Figdor, M. C., and Stern, C. D. (1993). Segmental organization of embryonic diencephalon. *Nature* **363:** 630–634.

Franks, R. G., and Crews, S. T. (1994). Transcriptional activation domains of the single-minded bHLH protein are required for CNS midline cell development. *Mech. Dev.* **45:** 269–277.

Frohman, M. A., Boyle, M., and Martin, G. R. (1990). Isolation of the mouse Hox-2.9 gene: Analysis of embryonic expression suggests that positional information along the anterior–posterior axis is specified by mesoderm. *Development* **110:** 589–607.

Green, E. L. (1981). Linkage, recombination, and mapping. In *Genetics and Probability in Animal Breeding Experiments* (E. L. Green, Ed.), pp. 77–83. Oxford Univ. Press, New York.

Green, M. C. (1989). Catalog of mutant genes and polymorphic loci. In *Genetic Variants and Strains of the Laboratory Mouse* (M. C. Green, Ed.), pp. 12–403. Oxford Univ. Press, New York.

Habets, G. G., van der Kammen, R. A., Jenkins, N. A., Gilbert, D. J., Copeland, N. G., Hagemeijer, A., and Collard, J. G. (1995). The invasion-inducing TIAM-1 gene maps to human chromosome 21q22 and mouse chromosome 16. *Cytogenet. Cell Genet.* **70:** 48–51.

Haefliger, J.-A., Bruzzone, R., Jenkins, N. A., Gilbert, D. J., Copeland, N. G., and Paul, D. L. (1992). Four novel members of the connexin family of gap junction proteins: Molecular cloning, expression and chromosome mapping. *J. Biol. Chem.* **267:** 2057–2064.

Heyman, S., Kent, A., and Lumsden, A. (1993). Cellular morphology and extracellular space at rhombomere boundaries in the chick embryo hindbrain. *Dev. Dynamics* **198:** 241–253.

Hoffman, E. C., Reyes, H., Chu, F. F., Sander, F., Conley, L. H., Brooks, B. A., and Hankinson, O. (1991). Cloning of a factor required for activity of the Ah (dioxin) receptor. *Science* **252:** 954–958.

Huang, Z. J., Edery, I., and Rosbash, M. (1993). PAS is a dimerization domain common to Drosophila period and several transcription factors. *Nature* **364:** 259–262.

Jain, S., Dolwick, K. M., Schmidt, J. V., and Bradfield, C. A. (1994). Potent transcription activation domains of the Ah receptor and the Ah receptor nuclear translocaror map to their carboxyl termini. *J. Biol. Chem.* **269:** 31518–31524.

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

Jenkins, N. A., Copeland, N. G., Taylor, B. A., and Lee, B. K. (1982). Organization, distribution, and stability of endogenous ecotropic murine leukemia virus DNA sequences in chromosomes of Mus musculus. *J. Virol.* **43:** 26–36.

Justice, M. J., Siracusa, L. D., Gilbert, D. J., Heisterkamp, N., Groffen, J., Chada, K., Silan, C. M., Copeland, N. G., and Jenkins, N. A. (1990). A genetic linkage map of mouse chromosome 10: Localization of 18 molecular markers using a single interspecific backcross. *Genetics* **125:** 855–866.

Klar, A., Baldassare, M., and Jessell, T. M. (1992). F-spondin: A gene expressed at high levels in the floor plate encodes a secreted protein that promotes neural cell adhesion and neurite extension. *Cell* **69:** 95–110.

Lewis, J. O., and Crews, S. T. (1994). Genetic analysis of the Drosophila single-minded gene reveals a central nervous system influence on muscle development. *Mech. Dev.* **48:** 81–91.

Lindebro, M. C., Poellinger, L., and Whitelaw, M. L. (1995). Protein–protein interaction via PAS domains: Role of the PAS domain in positive and negative regulation of the bHLH/PAS dioxin receptor-Arnt transcription factor complex. *EMBO J.* **14:** 3528–3539.

Maisonpierre, P., Belluscio, L., Squinto, S., Ip, N., Furth, M., Lindsay, R., and Yancopoulos, G. (1989). Neurotrophin-3: A neurotrophic factor related to NGF and BDNF. *Science* **247:** 1446–1451.

McMahon, A. P., Joyner, A. L., Bradley, A., and McMahon, J. A. (1992). The midbrain–hindbrain phenotype of Wnt-1-/Wnt-1- mice results from stepwise deletion of engrailed-expressing cells by 9.5 days postcoitum. *Cell* **69:** 581–595.

Muenke, M., Bone, L., Mitchell, H., Hart, I., Walton, K., Hall-Johnson, K., Ippel, E., Dietz-Band, J., Karloy, K., Fan, C.-M., Tessier-Lavigne, M., and Patterson, D. (1995). Physical mapping of the holoprosencephaly critical region in 21q22.3, exclusion of *SIM2* as a candidate gene for HPE, and mapping the *SIM2* to a region of chromosome 21 important for Down syndrome. *Am. J. Human Genet.* **57:** 1074–1079.

Nambu, J. R., Lewis, J. O., Wharton, K. J., and Crews, S. T. (1991). The Drosophila single-minded gene encodes a helix-loop-helix protein that acts as a master regulator of CNS midline development. *Cell* **67:** 1157–1167.

Olson, E. N., and Klein, W. H. (1994). bHLH factors in muscle development: Dead lines and commitments, what to leave in and what to leave out. *Genes Dev.* **8:** 1–8.

Ott, M.-O., Bober, E., Lyons, G., Arnold, H., and Buckingham, M. (1991). Early expression of the myogenic regulatory gene, *Myf5,* in precursor cells of skeletal muscle in the mouse embryo. *Development* **111:** 1097–1107.

Porteus, M. H., Bulfone, A., Ciaranello, R. D., and Rubenstein, J. L. R. (1991). Isolation and Characterization of a Novel cDNA Encoding a Homeodomain that is Developmentally Regulated in the Ventral Forebrain. *Neuron* **7:** 221–229.

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

Puelles, L., and Rubenstein, J. L. R. (1993). Expression patterns of homeobox and other putative regulatory genes in the embryonic mouse forebrain suggest a neuromeric organization. *Trends Neurosci.* **16:** 472–479.

Rao, V. N., Papas, T. S., and Reddy, E. S. (1987). erg, a human ets-related gene on chromosome 21: Alternative splicing, polyadenylation, and translation. *Science* **237:** 635–639.

Reisz-Porszasz, S., Probst, M. R., Fukunaga, B. N., and Hankinson, O. (1994). Identification of functional domains of the arcryl hydrocarbon receptor nuclear translocator protein (ARNT). *Mol. Cell. Biol.* **14:** 6075–6086.

Robinson, G. W., Wray, S., and Mahon, K. A. (1991). Spatially restricted expression of a member of a new family of murine Distal-less homeobox genes in the developing forebrain. *New Biol.* **3:** 1183–1194.

Rubenstein, J. L., Martinez, S., Shimamura, K., and Puelles, L. (1994). The embryonic vertebrate forebrain: The prosomeric model. *Science* **266:** 578–580.

Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989). *Molecular Cloning: A Laboratory Mannual,* 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Sanger, F., Nicklen, S., and Coulson, A. R. (1977). DNA sequencing with chain-termination inhibitors. *Proc. Natl. Acad. Sci. USA* **74:** 5463–5468.

Schaeren-Wiemers, N., and Gerfin-Moser, A. (1993). A single protocol to detect transcripts of various types and expression levels in neural tissue and cultured cells: in situ hybridization using digoxigenin-labelled cRNA probes. *Histochemistry* **100:** 431–440.

Shimamura, K., Hartigan, D. J., Matinez, S., Puelles, L., and Rubenstein, J. L. R. (1995). Longitudinal organization of the anterior neural plate and neural tube. *Development,* **121:** 3923–3933.

Wang, G. L., Jiang, B. H., Rue, E. A., and Semenza, G. L. (1995). Hypoxia-inducible factor 1 is a basic helix–loop helix–PAS heterodimer regulated by cellular O_2 tension. *Proc. Natl. Acad. Sci. USA* **92:** 5510–5514.

Weintraub, H., Davis, R., Tapscott, S., Thayer, M., Krause, M., Benezra, R., Blackwell, T. K., Turner, D., Rupp, R., Hollenberg, S., Zhuang, Y., and Lassar, A. (1991). The *myoD* gene family: Nodal point during specification of the muscle cell lineage. *Science* **248:** 761–765.

Whitelaw, M., Pongratz, I., Wilhelmsson, A., Gustafsson, J. A., and Poellinger, L. (1993). Ligand-dependent recruitment of the Arnt coregulator determines DNA recognition by the dioxin receptor. *Mol. Cell. Biol.* **13:** 2504–2514.

Whitelaw, M. L., Gottlicher, M., Gustafsson, J.-A., and Poellinger, L. (1993). Definition of a novel ligand binding domain of a nuclear bHLH receptor: Co-localization of ligand and HSP90 binding activities within the regulable inactivation domain of the dioxin receptor. *EMBO J.* **12:** 4169–4179.

Wilhelmsson, A., Cuthill, S., Denis, M., Wikstrom, A. C., Gustafsson, J. A., and Poellinger, L. (1990). The specific DNA binding activity of the dioxin receptor is modulated by the 90 kd heat shock protein. *EMBO J.* **9:** 69–76.

Wilkison, D. G., and Nieto, M. (1993). Detection of mesenger RNA by *in situ* hybrydization to tissue sections and whole mounts. *Methods Enzymol.* **225:** 361–373.

Wolf, C., Thisse, C., Stoetzel, C., Thisse, B., Gerlinger, P., and Perrin-Schmitt, F. (1991). The *M-twist* gene of *Mus* is expressed in subsets of mesodermal cells and is closely related to the *Xenopus X-twi* and the *Drosophila twist* genes. *Dev. Biol.* **143:** 363–373.

Yaginuma, H., Shiga, T., Homma, S., Ishihara, R., and Oppenheim, R. W. (1990). Identification of early developing axon projections from spinal interneurons in the chick embryo with a neuron specific β-tubulin antibody: Evidence for a new ''pioneer'' pathway in the spinal cord. *Development* **108:** 705–716.

Yamada, T., Placzek, M., Tanaka, H., Dodd, J., and Jessell, T. M. (1991). Control of cell pattern in the developing nervous system: polarizing activity of the floor plate and notochord. *Cell* **64:** 635–647.

Received December 15, 1995
Revised December 28, 1995
Accepted January 3, 1996