Cloning of Two Human Homologs of the *Drosophila single-minded* Gene SIM1 on Chromosome 6q and SIM2 on 21q Within the Down Syndrome Chromosomal Region

Roman Chrast,[1,2] Hamish S. Scott,[1] Haiming Chen,[1] Jun Kudoh,[4] Colette Rossier,[1] Shinsei Minoshima,[4] Yimin Wang,[4] Nobuyoshi Shimizu,[4] and Stylianos E. Antonarakis[1,2,3,5]

[1]Laboratory of Human Molecular Genetics, Department of Genetics and Microbiology and [2]Graduate Program of Cellular and Molecular Biology, Geneva University Medical School, 1211 Geneva, Switzerland; [3]Division of Medical Genetics, Cantonal Hospital of Geneva, 1211 Geneva, Switzerland; [4]Department of Molecular Biology, Keio University School of Medicine, Tokyo 160, Japan

As part of our effort to clone genes of human chromosome 21 that may contribute to Down syndrome, we have previously isolated four exons with homology to *Drosophila single-minded* (*sim*) gene, which encodes a transcription factor that is a master regulator of fruit fly neurogenesis. These exons were used to clone and characterize two human homologs of the *Drosophila sim* gene, SIM1 and SIM2, which map to chromosomes 6q16.3–q21 and 21q22.2, respectively; *SIM2* maps within the so-called Down syndrome chromosomal region. Recently, two mouse homologs, Sim1 and Sim2, also have been identified. There is a high level of homology among human, mouse, and *Drosophila sim* genes in their amino-terminal half where the conserved bHLH, PAS1, PAS2, and HST domains are present. In contrast, the carboxy-terminal parts are only homologous between SIM1 and Sim1 and SIM2 and Sim2. Two isoforms (SIM2 and SIM2s) of human SIM2 have been detected that differ in their 3′ ends. Northern blot analysis revealed one mRNA SIM1 species of ~9.5 kb and four different mRNA SIM2 species of 2.7, 3, 4.4, and 6 kb in human fetal kidney. The function of both human SIM1 and SIM2 is unknown. However, three copies of SIM2 may contribute to some specific Down syndrome phenotypes because of (1) mapping position, (2) potential function as transcriptional repressor, (3) likely dimerization with other transcription factors, (4) the temporal and spatial expression pattern of mouse Sim2, and (5) the potentially analogous role of human SIM2 to that of *Drosophila sim* during neurogenesis.

[The sequence data described in this paper have been submitted to GenBank under accession nos. U70212, U80456, U80457, and AB003185.]

Down syndrome (DS) is the most common autosomal aneuploidy occurring in 1.03 to 1.30 of 1000 live births (Epstein 1995). It is caused by the presence of three copies of human chromosome 21 (HC21). A minimum region of HC21 [Down syndrome chromosomal region (DSCR)] that in triplication is associated with many phenotypic characteristics of DS has been defined between D21S17 and ETS2 (McCormick et al. 1989; Delabar et al. 1993). The DSCR concept was challenged as some patients with DS phenotypes have triplicated re-

gions of HC21 proximal to this DSCR (Korenberg et al. 1994). However, the majority of partial trisomy 21 DS patients have three copies of the DSCR.

To identify genes that contribute to phenotypes of DS, we have performed exon trapping (Chen et al. 1996) on HC21 cosmids and identified a human homolog of the *Drosophila single-minded* gene (*sim*) that maps within the DSCR on HC21 (Chen et al. 1995). This localization subsequently was confirmed by others (Dahmane et al. 1995; Osoegawa et al. 1996).

The *Drosophila sim* gene is the master regulator of fruit fly neurogenesis (Thomas et al. 1988; Nambu et al. 1991). Sim protein is a transcription

[5]Corresponding author.
E-MAIL sea@medsun.unige.ch; FAX 41 22 702 57 06.

factor containing a basic helix–loop–helix (bHLH) motif, two PAS (PER/ARNT/SIM) domains, and the HST (HIF1-α/SIM/TRH) domain (Nambu et al. 1991; Isaac and Andrew 1996). During the course of this study, two mouse homologs of the *sim* gene (Sim1 and Sim2) were cloned (Ema et al. 1996a,b; Fan et al. 1996; Moffett et al. 1996; Yamaki et al. 1996). *Sim1* maps on mouse chromosome 10 and *Sim2* on mouse chromosome 16 in a region of synteny with HC21 (Fan et al. 1996). Both mouse Sim1 and Sim2 genes are expressed early [Sim2 from embryonic day 8.0 (E8.0) and Sim1 from E9.0] in developing forebrain (Ema et al. 1996b; Fan et al. 1996; Moffett et al. 1996; Yamaki et al. 1996) and outside the central nervous system (CNS), in somites, mesonephric duct, and foregut (Sim1), in facial and trunk cartilage, trunk muscles (Sim2), and in the developing kidney (Sim1 and Sim2) (Dahmane et al. 1995; Ema et al. 1996b; Fan et al. 1996; Moffett et al. 1996). In adult mouse, both Sim1 and Sim2 are expressed in kidney and skeletal muscles, whereas Sim2 is also expressed in the lung (Ema et al. 1996a,b; Moffett et al. 1996).

Here we describe the cloning of the cDNAs for two human homologs (SIM1 and SIM2) of the *Drosophila sim* gene, and mapping of *SIM1* to chromosome 6q16.3–q21 (human *SIM2* was mapped previously to 21q22.2; Chen et al. 1995). The function of the two mammalian SIM genes is unknown, but three copies of *SIM2* may contribute to some DS phenotypes.

RESULTS

Cloning of the Human SIM1 cDNA

Northern blot analysis using, as a probe, four trapped exons (see Methods) homologous to the *Drosophila sim* gene revealed expression only in human fetal kidney from the 16 adult and four fetal tissues tested (data not shown). A human fetal kidney cDNA library subsequently was screened using the same probe. The inserts of 14 positive plaques were partially sequenced; these sequences corresponded to two different SIM genes. Because during this study two mouse homologs of the *Drosophila sim* gene, Sim1 [GenBank accession nos. U40575 (Fan et al. 1996) and D79209 (Ema et al. 1996b)] and Sim2 [GenBank accession nos. U40576 (Fan et al. 1996), D63383 (Ema et al. 1996a), D64135 (Yamaki et al. 1996), and U42554 (Moffett et al. 1996)], were also described, we grouped the clones according to their homology to the corresponding mouse genes. Clone rch11 was 3.9 kb long and contained the

complete predicted coding region of a human homolog of the mouse Sim1 gene (Fig. 1). The human SIM1 (GenBank accession no. U70212) sequence predicted an open reading frame of 766 amino acids. The human SIM1 and mouse Sim1 genes showed a striking similarity. There is 87.4% nucleotide and 96.0% amino acid sequence identity between the SIM1 and Sim1 genes. The published mouse Sim1 sequence (GenBank accession no. U40575; Fan et al. 1996) contains an error in the translation termination codon region; the corrected sequence encodes for a longer predicted protein of 764 amino acids (T_{2332}GAAAAGGC to TGTTTTGGC; C.-M. Fan, pers. comm.). For our analysis we have used the mouse Sim1 (GenBank accession no. D79209) of Ema et al. (1996b). The initiation codon of SIM1 corresponded to that of the mouse Sim1 gene and conforms to the Kozak consensus sequence (Kozak 1987). The SIM1 protein (predicted molecular weight of 85 kD) contains a bHLH (amino acid residues 2–50), PAS1 (98–140), PAS2 (238–292), and HST (295–333) domains. The carboxyl-terminus contains a Ser-rich region (352–766; Fig. 2).

Mapping of SIM1 on Chromosome 6

Sim1 has been mapped to mouse chromosome 10 between markers Fyn and Ros1 (Fan et al. 1996). By synteny, the SIM1 gene was predicted to map on human chromosome 6 as confirmed by Muenke et al. (1995) using monochromosomal human–rodent hybrids. We have mapped by fluorescence in situ hybridization (FISH) *SIM1* on 6q16 using human bacterial artificial chromosome (BAC) KB153H1 containing the SIM1 gene as a probe (Fig. 3). To refine the localization of human *SIM1* gene we have used PCR [with oligonucleotides from the 3′ untranslated region (UTR) of SIM1 cDNA] of the Genebridge 4 radiation hybrid panel. The results of this analysis showed that *SIM1* mapped between markers WI-6516 and WI-6530 on 6q16.3–q21 [2.74 cR distal to WI-6516 (lod > 3.0)]. This interval is also defined by polymorphic markers D6S1041 and D6S468. Figure 4 summarizes the mapping positions of the mouse and human SIM genes.

Cloning of the Human SIM2 cDNA

The positive clones from the library screen that did not correspond to human SIM1 were homologous to the mouse Sim2 gene and were designated human SIM2. Sequencing of several clones (rch2, rch3, and rch10) revealed an ORF extending from codon
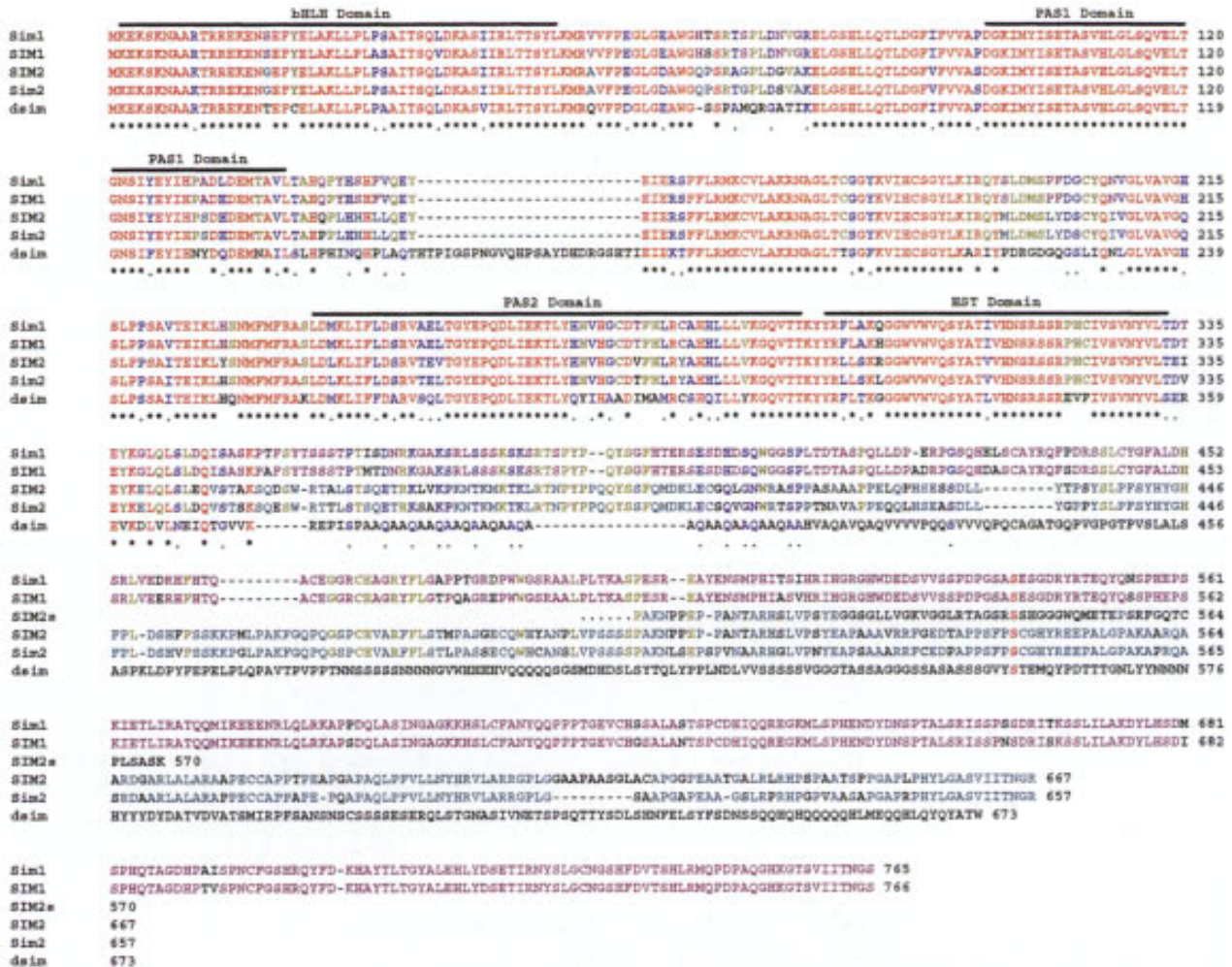
Figure 1 Comparison of human, mouse, and *Drosophila* SIM-predicted polypeptides. SIM1 (GenBank accession no. U70212) and SIM2 (U80456), human SIMs; Sim1 (D79209) and Sim2 (D63383), mouse SIMs; dsim (M19020), *Drosophila single minded*. SIM2s (U80457) represents the 570-amino-acid short isoform of SIM2. An asterisk (*) denotes identical and a period (.) denotes conserved amino acids, respectively. Amino acid color codes: (Red) Identical among all sequences; (blue) conserved among all sequences; (purple) identical between mouse and human SIM1; (turquoise) identical between mouse and human SIM2; (dark yellow) identical among all mammalian SIM proteins. SIM2s is shown from its point of divergence with SIM2.

4 to 525 of mouse Sim2. To obtain additional sequences from both ends of the human cDNA, 5′ and 3′ RACE were performed on human fetal kidney poly(A)⁺ mRNA. By homology to mouse Sim2, 5′ rapid amplification of cDNA ends (RACE) products completed the sequence of the 5′ coding region. The products obtained from the 3′ RACE did not differ from the 3′ ends of cDNA clones rch3 and rch10; all of these sequences were homologous to mouse Sim2 until codon 525 and then extended for an additional 1174 bp that ended with a poly(A) region and contained an additional in-phase ORF of 45 codons with no significant homologies. These sequences predicted a polypeptide of 570 amino acids, which we named SIM2 short form (SIM2s; GenBank accession no. U80457). A genomic sequence homologous to mouse Sim2 derived from the human BAC clone KB594G10 was used to determine additional sequences of SIM2 (SIM2; GenBank accession no. U80456) corresponding to the last 132 amino acids of mouse Sim2. Human BAC clone KB594G10 was mapped to HC21 by FISH (data not shown) and its 136-kb genomic sequence was determined completely (J. Kudoh, S. Minoshima, and N. Shimizu, in prep.). The predicted ORF of the long form of SIM2 was 667 amino acids long. The human
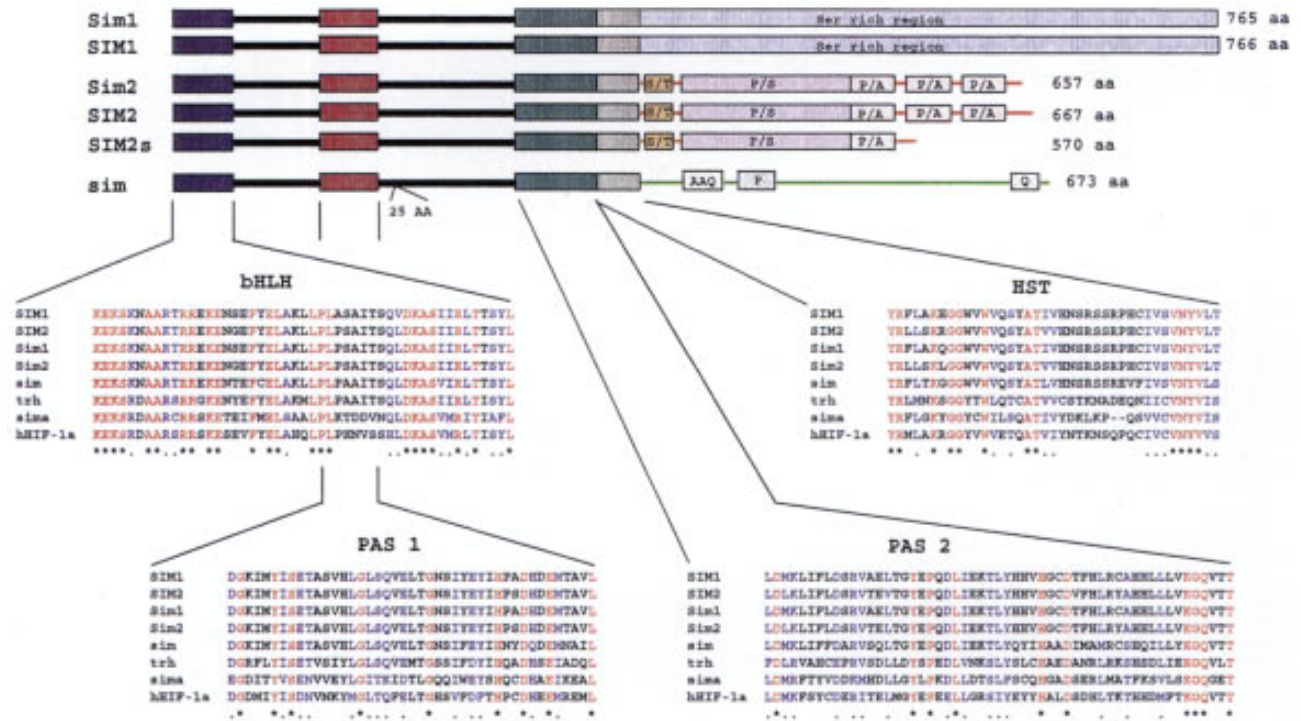
Figure 2 The conserved domains bHLH, PAS1, PAS2, and HST of the SIMs are shown schematically together with their alignments to domains from additional proteins. (Red) Identical amino acids; (blue) conserved amino acids. The GenBank numbers of the different SIM sequences used are as in Fig. 1; trh (U33427), *Drosophila trachealess;* sima (U43090), *Drosophila similar;* hHIF-1a (U22431), human hypoxia-inducible factor-1α. SIM2s represents the 570-amino-acid short isoform of SIM2. An asterisk (*) denotes identical and a period (.) denotes conserved amino acids.

SIM2 and mouse Sim2 genes are highly homologous; there is 86.7% nucleotide and 89.6% amino acid sequence identity between these two sequences. The translation initiation codon of SIM2 and SIM2s corresponds to that of mouse Sim2 and also conforms to the Kozak consensus (Kozak 1987). The SIM2 (predicted molecular mass 74-kD) and SIM2s proteins (predicted molecular mass 64-kD) both contain bHLH, PAS1, PAS2, and HST domains in the same position as described for SIM1 (Figs. 1 and 2; Table 1). The predicted carboxyl terminus of SIM2 and SIM2s both contain Ser/Thr-rich regions (from S348 to T366), Pro/Ser-rich regions (between P385 and S503), Pro/Ala-rich regions (from P504 to P544), and positively charged regions between R367 and R382. In addition, SIM2 contains two Pro/Ala-rich regions (from P553 to P596 and P611 to P644) and a positively charged region (from K559 to R575) (Fig. 2).

Northern Blot Analysis of Human SIM1 and SIM2

Northern blot analysis using four trapped exons (see Methods) from the conserved region between SIM1 and SIM2 genes revealed five hybridizing mRNA species of ~9.5, 6, 4.4, 3, and 2.7 kb only in fetal kidney from 20 different human adult and fetal tissues tested (data not shown). Hybridization using the 3′ half of SIM1 between nucleotides 1364 and 1610 (which is not homologous to SIM2), revealed only the 9.5-kb mRNA species in fetal kidney (Fig. 5, left). In contrast, a SIM2-specific probe (from nucleotide 1140 to 1433) hybridized to the mRNA species of ~6, 4.4, 3, and 2.7 kb (Fig. 5, right). The 9.5-kb mRNA is therefore specific for SIM1, and the other four mRNA species are specific for SIM2. The different SIM2-hybridizing mRNAs could be attributable to different 5′ or 3′ UTRs, alternative splicing, and/or overlapping transcripts.

The Conserved Domains in Human, Mouse, and *Drosophila sim* Genes

All five cDNAs (*Drosophila sim*, mouse Sim1 and Sim2, and human SIM1 and SIM2) and their predicted polypeptides are remarkably similar to each
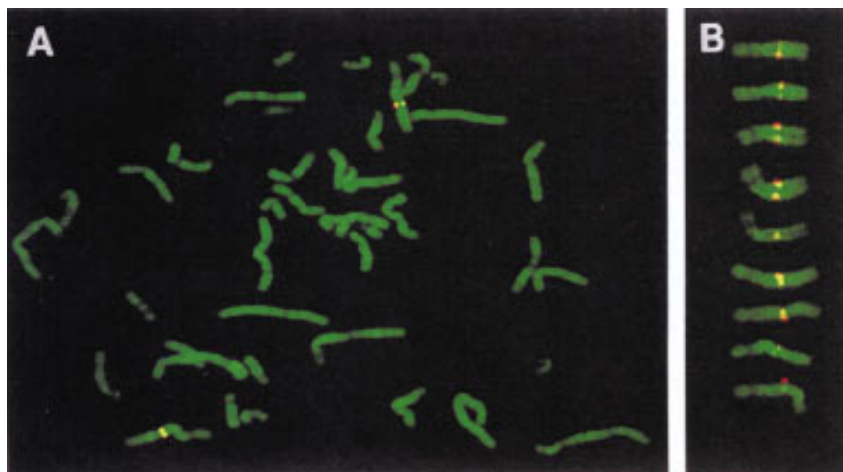
Figure 3 Chromosomal localization of the BAC clone KB153H1 containing the human SIM1 gene to 6q16 by FISH and Q-banding. (*A*) Metaphase chromosomes of human B-lymphoblastoid GM130B cell line hybridized with the BAC clone KB153H1. (*B*) Representative images of human chromosome 6. Hybridization signals appeared in the same position, 6q16, in all cases.

mologs of the *Drosophila sim* gene that is involved in neuronal development. We have mapped *SIM1* to 6q16.3–q21, whereas *SIM2* maps to 21q22.2. The two mouse homologs of *sim* have also been described: *Sim1* (Ema et al. 1996b; Fan et al. 1996) and *Sim2* (Ema et al. 1996a; Fan et al. 1996; Moffett et al. 1996; Yamaki et al. 1996), which have 96% and 89.6% amino acid sequence identity to human SIM1 and SIM2, respectively. cDNA library screening (fetal kidney) and 3′ RACE (on fetal kidney mRNA) resulted only in a short isoform of SIM2 (SIM2s). Genomic sequencing revealed a longer form of SIM2 with homology to the entire mouse Sim2 cDNA. RT–PCR using oligonucleotides derived from the sequences of the long form failed to amplify the expected long form product from fetal kidney mRNA. However, this result is consistent with the observation that the sequence corresponding to the carboxy-terminal coding region of SIM2 gene was difficult to amplify by PCR, even from cloned DNA fragments. The short isoform, SIM2s, is probably attributable to the use of alternative 3′ UTR contained within intron 10 of SIM2 or mispriming on an A-rich sequence within this intron. The carboxyl terminus of SIM2 (from A526 to R667) is encoded by exon 11, which is separated from exon 10 by a 2528-bp intron 10. The nucleotide sequence encoding the carboxyl terminus of SIM2s (from G526 to K570) starts from the last nucleotide of exon 10 and continues to the sequence of intron 10 (GenBank accession no. AB003185; J. Kudoh, S. Minoshima, and N. Shimizu, in prep.). Whether SIM2s is indeed a functional isoform in different tissues and stages of development remains unknown.

other in the amino terminus half up to the amino acid corresponding to K351 of SIM1 (Fig. 1). Table 1 displays the percent identity among different regions of the sequences examined. The amino acid sequence identity among the five predicted polypeptides for their first half is between 71% and 99%. Within this region of remarkable homology, all of these proteins contain four recognized domains, namely bHLH, PAS1, PAS2, and HST, often seen in transcription factors (Fig. 1). Figure 2 shows a partial amino acid alignment of proteins that contain all four of these domains. These proteins, in addition to the five already mentioned, include the *Drosophila trachealess* (*trh*; Isaac and Andrew 1996), *Drosophila similar* (*sima*; Nambu et al. 1996), and human hypoxia-inducible factor 1α (HIF-1α; Wang et al. 1995). Within the bHLH domain, all eight proteins have 69% identical and conserved amino acids. These percentages for the PAS1, PAS2, and HST domains are 53%, 45%, and 56%, respectively.

In the carboxy-terminal parts of the predicted amino acid sequences, there are no apparent homologies between *Drosophila sim* and its mammalian homologs (Fig. 1). However, SIM1 and SIM2 showed high conservation of sequences in the carboxy-terminal part when compared with mouse Sim1 and Sim2, respectively (Table 1; Figs. 1 and 2).

DISCUSSION

We have isolated the cDNAs of two human ho-

amplify the expected long form product from fetal kidney mRNA. However, this result is consistent with the observation that the sequence corresponding to the carboxy-terminal coding region of SIM2 gene was difficult to amplify by PCR, even from cloned DNA fragments. The short isoform, SIM2s, is probably attributable to the use of alternative 3′ UTR contained within intron 10 of SIM2 or mispriming on an A-rich sequence within this intron. The carboxyl terminus of SIM2 (from A526 to R667) is encoded by exon 11, which is separated from exon 10 by a 2528-bp intron 10. The nucleotide sequence encoding the carboxyl terminus of SIM2s (from G526 to K570) starts from the last nucleotide of exon 10 and continues to the sequence of intron 10 (GenBank accession no. AB003185; J. Kudoh, S. Minoshima, and N. Shimizu, in prep.). Whether SIM2s is indeed a functional isoform in different tissues and stages of development remains unknown.

Each of the human, murine, and *Drosophila* SIM genes contains highly conserved bHLH, PAS1, PAS2, and HST domains. The bHLH domain, present in many transcription factors, including *myc*, MyoD, and E2A (Murre et al. 1989), contains two amphipathic α helices separated by a variable loop, which participate in the formation of dimers, and a basic region that mediates DNA binding (Murre et al. 1989; Davis et al. 1990). The PAS1 and PAS2 domains were first described in the *Drosophila sim* gene (Nambu et al. 1991) and are present in other pro-
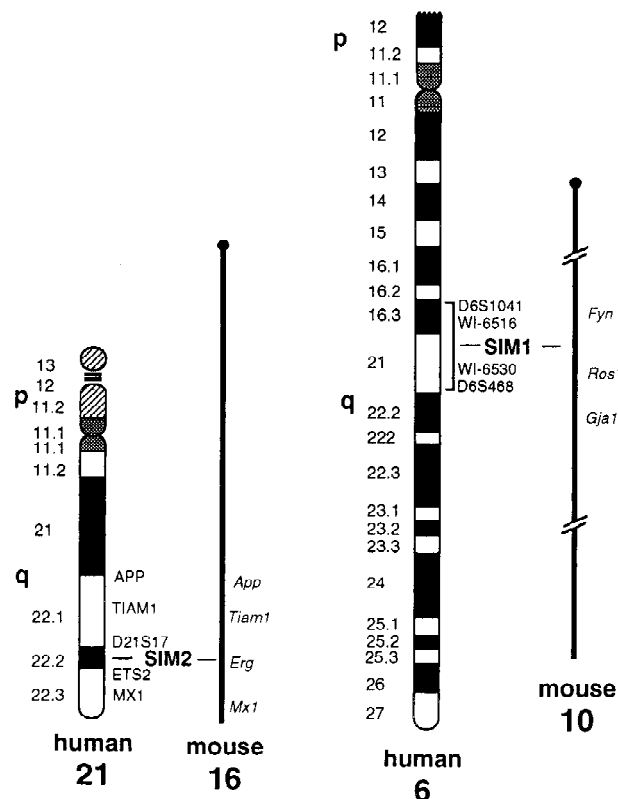
Figure 4 Mapping position of human and mouse *SIM1* and *SIM2* genes. The localization in mouse chromosomes was reported in Fan et al. (1996). Previously we have mapped *SIM2* to chromosome 21 (Chen et al. 1995). *SIM1* has been mapped to chromosome 6q16.3–q21 between markers WI-6516 and WI-6530 using the Genebridge 4 radiation hybrid panel. Diagrams are not to scale.

the known PAS-containing proteins (Arnt, Arnt2, Ahr, and HIF1-α) have been tested for heterodimerization with Sim. Ema et al. (1996b) also showed that Sim1 and Sim2 do not form homodimers or dimerize with each other. The carboxy-terminal parts of mouse Sim1 and Sim2 are different from those of other bHLH/PAS family members (Arnt, Arnt2, Ahr, HIF-1α, and *sim*); this can explain the fact that both Sim1 and Sim2 function as transcriptional repressors, whereas all of the other proteins mentioned function as transcriptional activators (Franks and Crews 1994; Jain et al. 1994; Sogawa et al. 1995a; Hirose et al. 1996). No specific domain in the carboxy-terminal half of SIM1 (from amino acid 352) has been recognized, except that this region is rich in serine residues, as in mouse Sim1 (Ema et al. 1996b). In the carboxy-terminal half of SIM2 there are regions similar to those in mouse Sim2 (Yamaki et al. 1996) rich in serine/threonine, proline/serine, and proline/alanine. Domains rich in serine/ threonine and proline residues are present in both transcription repressors (Madden et al. 1991) and transcription activators (Mermod et al. 1989; Franks and Crews 1994). Based on the amino acid homologies of human SIM1 and SIM2 to their mouse counterparts it is logical to predict that they also could function as transcriptional repressors.

Because the carboxy-terminal half of *Drosophila sim* is not homologous to those found in human SIM1/SIM2 and mouse Sim1/Sim2, it is possible that these vertebrate homologs of the *Drosophila sim* gene do not represent its true homologs. There may be another vertebrate SIM gene with homology spanning the 3′ half or the *Drosophila sim,* although such mammalian expressed sequence tags (ESTs) have not been found. It is also possible that the *Drosophila sim* transcriptional activation domains in the carboxyl terminus had changed during evolution and this region of the protein had acquired repression function.

Phenotypes that may be associated with deleterious mutations in *SIM1* and *SIM2* are unknown. No candidate Mendelian or complex phenotype has been mapped to the chromosomal localizations of the human and mouse single-minded homologs. Mice with targeted disruption of either *Sim1* or *Sim2* may contribute to the elucidation of their involvement in normal mammalian development and pathological conditions associated with null mutations. The study of the phenotype of mice heterozygous for a disrupted Sim2 gene will potentially elucidate the contribution of this gene in partial monosomy 21 that includes 21q22.2 (Huret et al. 1995).

teins including (1) the *Drosophila period* (*per*) (Hardin et al. 1990); (2) the vertebrate aryl hydrocarbon (Ah) receptor (AhR) and (3) AhR nuclear translocators (Arnt and Arnt2) (Hoffman et al. 1991; Reyes et al. 1992; Hirose et al. 1996); (4) the human HIF-1α (Wang et al. 1995); (5) the *Drosophila trh* (Isaac and Andrew 1996), and (6) the *Drosophila sima* (Nambu et al. 1996). PAS domains are involved in the dimerization of these proteins (Huang et al. 1993; Sogawa et al. 1995b). The HST domain is conserved among all sim genes, *trh, sima,* and HIF-1α; its function is presently unknown.

It is likely that both human SIM proteins are transcription factors that form dimers because they contain the domains described above. Both mouse Sim proteins can heterodimerize with Arnt and Arnt2 (Ema et al. 1996b); however, it is possible that there are additional partners for SIM because only

Table 1. Percent Identity among Human, Mouse, and *Drosophila* SIMs

	SIM2s		SIM2		Sim1		Sim2		sim	
	nt	aa	nt	aa	nt	aa	nt	aa	nt	aa
SIM1 (1–351)	74.4	87.2	74.4	87.2	88.6	98.9	76	88.9	58.3	71.8
SIM2 (1–351)					72.6	87.7	89.5	96.9	57	70.9
Sim1 (1–351)							75.1	89.4	55	72.1
Sim2 (1–351)									56.1	71.2
SIM1 (352–end)	28.4	19.6	22.9	18	86.2	93.7	22.8	16	20.8	11.7
SIM2s (352–end)			82.1	81.3	25.5	17.3	69.2	68	20.1	11.4
SIM2 (352–end)					22.2	15.8	82.6	81.4	19.8	9.7
Sim1 (352–end)							21.5	15.3	21.4	10.7
Sim2 (352–end)									20.1	9.4

Percent identity on the nucleotide (nt) and amino acid (aa) level (calculated using ClustalW 1.6). The GenBank accession numbers of the different SIM sequences used are as in Fig. 1.

SIM2, with its three copies, is an exceptional candidate to be associated with certain phenotypes of DS because (1) it maps to the chromosomal region that is present in three copies in the majority of DS individuals with partial trisomy 21 (Delabar et al. 1993); (2) the *Drosophila sim* is a master regulator of fruit fly neurogenesis and mutations in the *sim* gene result in loss of precursor cells that give rise to the midline cells of the embryonic CNS (Crews et al.
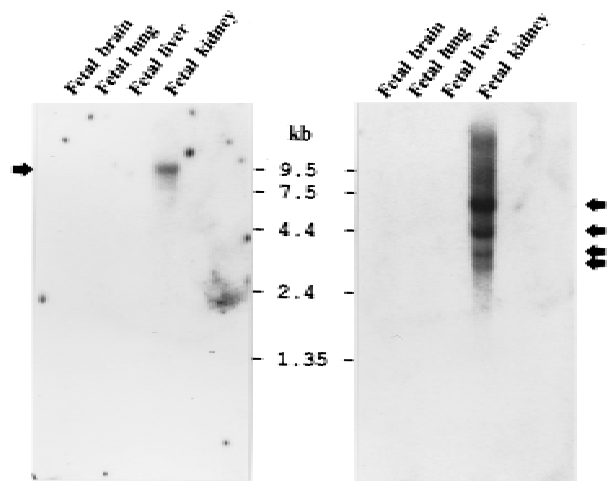

Figure 5 Northern blot analysis of SIM1 and SIM2 expression. The nonhomologous carboxy-terminal parts of SIM1 and SIM2 cDNAs were used as specific probes against filters containing poly(A)$^+$ RNA from several human tissues (see Methods). The same filters were used sequentially for both hybridizations. Both genes are expressed in fetal kidney; SIM1 revealed an mRNA species of ~9.5 kb (*left*), whereas SIM2 showed mRNA species of ~2.7, 3, 4.4, and 6 kb (*right*).

1988; Nambu et al. 1991); (3) the mouse Sim2 gene is expressed early (E8.0) in the forebrain of mouse embryo (Ema et al. 1996b; Fan et al. 1996; Moffett et al. 1996; Yamaki et al. 1996). It has been shown that DS patients have selective anatomic abnormalities in brain compared with normals (Raz et al. 1995; Kesslak et al. 1994). Outside the CNS, Sim2 is expressed in facial and trunk cartilage, trunk muscles, and kidney (Dahmane et al. 1995; Ema et al. 1996b; Fan et al. 1996; Moffett et al. 1996). We have confirmed these data using whole-mount in situ hybridization of E11.5 mouse embryos with an antisense Sim2-specific probe (R. Chrast, Y. Herault, S. Antonarakis, and D. Duboule, unpubl.). In addition, (4) SIM2 is a potential transcriptional repressor, and (5) it may form heterodimers. Its overexpression may alter the kinetics of dimer formation resulting in abnormal regulation of downstream genes. The generation of Sim2 transgenic mice may enhance our understanding of the involvement of this gene in DS. A mouse model for DS, attributable to partial mouse trisomy 16, has been reported (Reeves et al. 1995). This model contains three copies of Sim2, among many other genes. Mating of this animal with a Sim2 knockout mouse would result in partial trisomy 16 mice with two copies of Sim2 and thus may elucidate the contribution of Sim2 to the phenotype in the context of this partial trisomy 16.

METHODS

Cloning of Human SIM1 and SIM2 cDNAs

To clone the full-length coding sequences of the human SIM1 and SIM2 cDNAs ~10^6 PFUs of a human fetal kidney 5′ stretch

λDR2 cDNA library (Clontech, HL1150x) was screened using a probe derived from a mixture of four previously trapped exons homologous to the *Drosophila sim* gene (HMC13F06, HMC29C01, HMC05F04, and RCH03D09; GenBank accession nos. X83514, X83515, X83513, and X83516, respectively) (Chen et al. 1995). Radioactive ^{32}P-labeling by PCR was performed as described (Scott et al. 1995), except that the exon-trapping vector pSPL3 (GIBCO BRL, 18449-017)-derived oligonucleotides dUSD2 (5′-CUACUACUACUAGTGAACTG-CACTGTGACAAGCTGC-3′) and dUSA4 (5′-CUACUACU-ACUACACCTGAGGAGTGAATTGGTCG-3′) were used for PCR amplification. The inserts of positive phages were converted to plasmids (pDR2) as described (Clontech, PT1011-1) and sequenced using the ABI373A sequencer. The sequence contigs of SIM1 and SIM2 cDNAs were made using the SeqEd program (ABI); nucleotide and predicted amino acid homologies were analyzed by use of BLAST (Altschul et al. 1990). Analyses of the resulting sequences were performed with numerous programs of the Tools option of EXPASY (http://expasy.hcuge.ch/www/tools.html). Multiple sequence alignments were made with ClustalW 1.6 (http://dot.imgen.bcm.tmc.edu:9331/multi-align/multi-align.html).

To obtain 5′ and 3′ coding sequences of SIM2 cDNA, RACE was performed on fetal kidney poly(A)$^+$ mRNA (Clontech, 6526-1) using a Marathon cDNA amplification kit (Clontech, K1802-1). Double-stranded cDNA synthesis and adapter ligation to the synthesized cDNAs were carried out according to manufacturer's protocol PT1115-1. A first-round RT–PCR amplification was performed on the synthesized cDNA with primers AP1 (see protocol) and SM2RE2 (5′-AGTGCGATCCCAGCTCCTTGGCGA-3′) for the 5′ RACE or AP1 and SM2F03 (5′-GCACAATGCCAGCCAGCGGTGAAT-3′) for the 3′ RACE. A nested PCR amplification was then performed with primers AP2 (see protocol) and SM2RE1 (5′-GAAGACGGCGCGCATCTTCAGGTA-3′) for the 5′ RACE or AP2 and SM2F04 (5′-GAATGCCAGTGGCATTATGCCAAC-3′) for the 3′ RACE. The PCR products have been purified using QIAquick PCR purification kit (Qiagen) and subjected to sequencing.

Only the short form of SIM2 (SIM2s) has been found by cDNA library screening or 3′ RACE; the sequence of the long form of SIM2, homologous to mouse Sim2, was determined by genomic sequencing of the appropriate fragment of human BAC clone KB594G10 as described (Kawasaki et al. 1997).

Northern Blot Analysis

PCR-derived probes of 246 and 293 bp from the 3′ half of SIM1 and SIM2 cDNAs, respectively, which showed no homology to each other, were prepared as described (Scott et al. 1995) using oligonucleotides S11D09F3 5′-GCTGGTG-GAAGAGAGGCATT-3′ and S11D09R4 5′-TGGAGAACTGAC-CACACTAT-3′ for SIM1, and S2F04F2 5′-GCTGAGAACAAAC-CCTTACC-3′ and S2F04R3 5′-AAGCGTGCCACCTCACAA-3′ for SIM2. The human adult and fetal multiple tissue Northern blots (Clontech, 7756-1, 7760-1, and 7759-1) were hybridized and washed at high stringency.

Mapping of SIM1

The Genebridge 4 radiation hybrid DNA panel (Walter et al. 1994) (HGMP resource center; http://www.hgmp.mrc.ac.uk/Public/Docs/Bio/Hybrid_panels.html) was tested by PCR for the presence of the specific 115-bp amplification fragment derived from the 3′ UTR of the SIM1 gene by use of oligonucleotides SIM1F 5′-TTCATGTAGATCAAGTGTGC-3′ and SIM1R 5′-ACCGCAGTTTGTAAGTAATA-3′. The scoring was performed on a 1.2% agarose gel and confirmed by hybridization. The data were submitted to and analyzed by the MIT Radiation Hybrid Mapping site (http://www-genome.wi.mit.edu/cgi-bin/contig/rhmapper.pl).

FISH was performed on Q-banded metaphase chromosomes of human B-lymphoblastoid cell line GM00130B as described (Wang et al. 1996) using BAC clone KB153H1 containing the human SIM1 gene as a probe. Human BAC clones KB153H1, containing the SIM1 gene, and KB594G10, containing the SIM2 gene, were isolated from a total human BAC library using a mouse Sim2 cDNA clone, mSIMλ1 (Yamaki et al. 1996), as a probe (Asakawa et al. 1997).

REFERENCES

Altschul, S.F., W. Gish, W. Miller, E.W. Myers, and D.J. Lipman. 1990. Basic local alignment search tool. *J. Mol. Biol.* **215:** 403–410.

Asakawa, S., I. Abe, Y. Kudoh, N. Kishi, Y. Wang, R. Kubota, J. Kodoh, K. Kawasaki, S. Minoshima, and N. Shimizu. 1997. Human BAC library: Construction and rapid screening. *Gene* (in press).

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

Chen, H., R. Chrast, C. Rossier, M.A. Morris, M.D. Lalioti, and S.E. Antonarakis. 1996. Cloning of 559 potential exons of genes of human chromosome 21 by exon trapping. *Genome Res.* **6:** 747–760.

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

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

Davis, R.L., P.-F. Cheng, A.B. Lassar, and H. Weintraub. 1990. The MyoD DNA binding domain contains a recognition code for muscle-specific gene activation. *Cell* **60:** 733–746.

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

Ema, M., M. Suzuki, M. Morita, K. Hirose, K. Sogawa, Y. Matsuda, O. Gotoh, Y. Saijoh, H. Fujii, H. Hamada, and Y. Fujii-Kuriyama. 1996a. cDNA cloning of a murine homologue of Drosophila single-minded, its mRNA expression in mouse development, and chromosome localization. *Biochem. Biophys. Res. Commun.* **218:** 588–594.

Ema, M., M. Morita, S. Ikawa, M. Tanaka, Y. Matsuda, O. Gotoh, Y. Saijoh, H. Fujii, H. Hamada, Y. Kikuchi, and Y. Fujii-Kuriyama. 1996b. Two new members of the murine *Sim* gene family are transcriptional repressors and show different expression patterns during mouse embryogenesis. *Mol. Cell. Biol.* **16:** 5865–5875.

Epstein, C.D. 1995. Down syndrome (Trisomy 21). In *The metabolic and molecular basis of inherited disease,* vol. I, 7th ed. (ed. C.R. Scriver, A.L. Beaudet, W.S. Sly, and D. Valle), pp. 749–794. McGraw-Hill, New York, NY.

Fan, C.-M., E. Kuwana, A. Bulfone, C.F. Fletcher, N.G. Copeland, N.A. Jenkins, S. Crews, S. Martinez, L. Puelles, J.L.R. Rubenstein, and M. Tessier-Lavigne. 1996. Expression patterns of two murine homologs of Drosophila single-minded suggest possible roles in embryonic patterning and in the pathogenesis of Down syndrome. *Mol. Cell. Neurosci.* **7:** 1–16.

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

Hardin, P.E., J.C. Hall, and M. Rosbash. 1990. Feedback of the *Drosophila period* gene product on circadian cycling of its messenger RNA levels. *Nature* **343:** 536–540.

Hirose, K., M. Morita, M. Ema, J. Mimura, H. Hamada, H. Fujii, Y. Saijo, O. Gotoh, K. Sogawa, and Y. Fujii-Kuriyama. 1996. cDNA cloning and tissue-specific expression of a novel bHLH/PAS factor (Arnt2) with close sequence similarity to AhR nuclear translocator (Arnt). *Mol. Cell. Biol.* **16:** 1706–1713.

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

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

Huret, J.-L., C. Leonard, M. Chery, C. Philippe, E. Shafei-Benaissa, G. Lefaure, B. Labrune, and S. Gilgenkrantz. 1995. Monosomy 21q: Two cases of del(21q) and review of the literature. *Clin. Genet.* **48:** 140–147.

Isaac, D.D. and D.J. Andrew. 1996. Tubulogenesis in *Drosophila*: A requirement for the trachealess gene product. *Genes & Dev.* **10:** 103–117.

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

Kawasaki, K., S. Minoshima, E. Nakato, K. Shibuya, A. Shintani, J.L. Schmeits, J. Wang, and N. Shimizu. 1997. One megabase sequence analysis of the human immunoglobulin λ gene locus. *Genome Res.* **7:** 250–261.

Kesslak, J.P., S.F. Nagata, I. Lott, and O. Nalcioglu. 1994. Magnetic resonance imaging analysis of age-related changes in the brains of individuals with Down's syndrome. *Neurology* **44:** 1039–1045.

Korenberg, J.R., X.N. Chen, R. Schipper, Z. Sun, R. Gonsky, S. Gerwehr, N. Carpenter, C. Daumer, P. Dignan, C. Disteche, J.M. Graham, L. Hugdins, B. McGillivray, K. Miyazaki, N. Ogasawara, J.P. Park, R. Pagon, S. Pueschel, G. Sack, B. Say, S. Schuffenhauer, S. Soukup, and T. Yamanaka. 1994. Down syndrome phenotypes: The consequences of chromosomal imbalance. *Proc. Natl. Acad. Sci.* **91:** 4997–5001.

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

Madden, S.L., D.M. Cook, J.F. Morris, A. Gashler, V.P. Sukhatme, and F.J. Rauscher III. 1991. Transcriptional repression mediated by the WT1 Wilms tumour gene product. *Science* **253:** 1550–1553.

McCormick, M.K., A. Schinzel, M.B. Petersen, G. Stetten, D.J. Driscoll, E.S. Cantu, L. Tranebjaerg, M. Mikkelsen, P.C. Watkins, and S.E. Antonarakis. 1989. Molecular genetic approach to the characterization of the ''Down syndrome region'' of chromosome 21. *Genomics* **5:** 325–331.

Mermod, N., E.A. O'Neill, T.J. Kelly, and R. Tjian. 1989. The proline-rich transcriptional activator of CTF/NF-I is distinct from the replication and DNA binding domain. *Cell* **58:** 741–753.

Moffett, P., M. Dayo, M. Reece, M.K. McCormick, and J. Pelletier. 1996. Characterization of *msim,* a murine homolog

of the *Drosophila sim* transcription factor. *Genomics* **35:** 144–155.

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

Murre, C., P.S. McCaw, H. Vaessin, M. Caudy, L.Y. Jan, Y.N. Jan, C.V. Cabrera, J.N. Buskin, S.D. Hauschka, A.B. Lassar, H. Weintraub, and D. Baltimore. 1989. Interactions between heterologous helix–loop–helix proteins generate complexes that bind specifically to a common DNA sequence. *Cell* **58:** 537–544.

Nambu, J.R., J.O. Lewis, K.A. Wharton Jr, and S.T. Crews. 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.

Nambu, J.R., W. Chen, S. Hu, and S.T. Crews. 1996. The *Drosophila melanogaster similar* bHLH–PAS gene encodes a protein related to human hypoxia-inducible factor 1α and Drosophila single-minded. *Gene* **172:** 249–254.

Osoegawa, K., S. Okano, Y. Kato, Y. Nishimura, and E. Soeda. 1996. A 19-kb CpG island associated with single-minded gene 2 in Down syndrome chromosomal region. *DNA Res.* **3:** 175–179.

Raz, N., I.J. Torres, S.D. Briggs, W.D. Spencer, A.E. Thornton, W.J. Loken, F.M. Gunning, J.D. McQuain, N.R. Driesen, and J.D. Acker. 1995. Selective neuroanatomic abnormalities in Down's syndrome and their cognitive correlates: Evidence from MRI morphometry. *Neurology* **45:** 356–366.

Reeves, R.H., N.G. Irving, T.H. Moran, A. Wohn, C. Kitt, S.S. Sisodia, C. Schmidt, R.T. Bronson, and M.T. Davisson. 1995. A mouse model for Down syndrome exhibits learning and behaviour deficits. *Nat. Genet.* **11:** 177–184.

Reyes, H., S. Reisz-Porszasz, and O. Hankinson. 1992. Identification of the Ah receptor nuclear translocator protein (Arnt) as a component of the DNA binding form of the Ah receptor. *Science* **256:** 1193–1195.

Scott, H.S., L. Blanch, X.H. Guo, C. Freeman, A. Orsborn, E. Baker, G.R. Sutherland, C.P. Morris, and J.J. Hopwood. 1995. Cloning of the sulphamidase gene and identification of mutations in Sanfilippo A syndrome. *Nature Genet.* **11:** 465–467.

Sogawa, K., K. Iwabuchi, H. Abe, and Y. Fujii-Kuriyama. 1995a. Transcriptional activation domains of the Ah receptor and Ah receptor nuclear translocator. *J. Cancer Res. Clin. Oncol.* **121:** 612–620.

Sogawa, K., R. Nakano, A. Kobayashi, Y. Kikuchi, N. Ohe, N. Matsushita, and Y. Fujii-Kuriyama. 1995b. Possible function of Ah receptor nuclear translocator (Arnt) homodimer in transcriptional regulation. *Proc. Natl. Acad. Sci.* **92:** 1936–1940.

Thomas, J.B., S.T. Crews, and C.S. Goodman. 1988. Molecular genetics of the single-minded locus: A gene involved in the development of the drosophila nervous system. *Cell* **52:** 133–141.

Walter, M.A., D.J. Spillett, P. Thomas, J. Weissenbach, and P.N. Goodfellow. 1994. A method for constructing radiation hybrid maps of whole genomes. *Nature Genet.* **7:** 22–28.

Wang, G.L., B.H. Jiang, E.A. Rue, and G.L. Semenza. 1995. Hypoxia-inducible factor 1 is a basic–helix–loop–helix–PAS heterodimer regulated by cellular O2 tension. *Proc. Natl. Acad. Sci.* **92:** 5510–5514.

Wang, Y., J. Kudoh, R. Kubota, S. Asakawa, S. Minoshima, and N. Shimizu. 1996. Chromosomal mapping of a family of human glutamine synthetase genes: Functional gene (GLUL) on 1q25, pseudogene (GLULP) on 9p13, and three related genes (GLULL1, GLULL2, GLULL3) on 5q33, 11p15, and 11p24. *Genomics* **37:** 195–199.

Yamaki, A., S. Noda, J. Kudoh, N. Shindoh, H. Maeda, S. Minoshima, K. Kawasaki, Y. Shimizu, and N. Shimizu. 1996. The mammalian *Single-minded* (*SIM*) gene: Mouse cDNA structure and diencephalic expression indicate a candidate gene for Down syndrome. *Genomics* **35:** 136–143.