Published in final edited form as:
Biochim Biophys Acta. 2010 July ; 1799(7): 520–531. doi:10.1016/j.bbagrm.2010.03.002.

Transcription of mouse *Sp2* yields alternatively spliced and sub-genomic mRNAs in a tissue- and cell type-specific fashion

Haifeng Yin, **Teresa D. Nichols**, and **Jonathan M. Horowitz***
Department of Molecular Biomedical Sciences and the Center for Comparative Medicine and Translational Research, College of Veterinary Medicine, North Carolina State University, Raleigh, NC 27606, USA

Abstract

The Sp-family of transcription factors is comprised by nine members, *Sp1-9*, that share a highly-conserved DNA-binding domain. *Sp2* is a poorly characterized member of this transcription factor family that is widely expressed in murine and human cell lines yet exhibits little DNA-binding or *trans*-activation activity in these settings. As a prelude to the generation of a "knock-out" mouse strain, we isolated a mouse *Sp2* cDNA and performed a detailed analysis of *Sp2* transcription in embryonic and adult mouse tissues. We report that (1) the 5′ untranslated region of *Sp2* is subject to alternative splicing, (2) *Sp2* transcription is regulated by at least two promoters that differ in their cell-type specificity, (3) one *Sp2* promoter is highly active in nine mammalian cell lines and strains and is regulated by at least five discrete stimulatory and inhibitory elements, (4) a variety of sub-genomic messages are synthesized from the *Sp2* locus in a tissue- and cell type-specific fashion and these transcripts have the capacity to encode a novel partial-Sp2 protein, and (5) RNA *in situ* hybridization assays indicate that *Sp2* is widely expressed during mouse embryogenesis, particularly in the embryonic brain, and robust *Sp2* expression occurs in neurogenic regions of the post-natal and adult brain.

Keywords

Sp2; Promoter analysis; Exonic promoter; Alternative splicing; *In situ* hybridization

1. Introduction

The largest family of sequence-specific DNA-binding proteins in the human proteome carries DNA-binding domains that feature one or more zinc-"fingers" [1,2]. The Sp family of transcription factors form one sub-class of these proteins. Nine members of the Sp family have been identified and each shares two common structural features: a highly conserved carboxy-terminal DNA-binding domain and an amino-terminal *trans*-activation domain. Four conserved nucleotides within each zinc-"finger" specify the interaction of the DNA-binding domain with a nonameric GC-rich sequence that is often located proximal to sites of transcriptional initiation. A subset of Sp-family members share additional sequences of functional significance within their respective *trans*-activation domains. The *trans*-activation

* Corresponding author: 4700 Hillsborough Street, Raleigh, NC 27606. Phone: 919-515-4479; Fax: 919-513-7301. jon_horowitz@ncsu.edu (J. M. Horowitz).

domains of Sp1-Sp4 are comprised of three sub-domains, termed A, B, and C. Domains A and B each contain a serine- and threonine-rich region ($A^{S/T}$, $B^{S/T}$) followed by a glutamine-rich region (A^{Q}, B^{Q}; [3,4]). The C domain encodes the most carboxy-terminal portion of the *trans*-activation domain, features a preponderance of charged amino acids, and carries a portion of a bipartite nuclear localization sequence [5,6]. Sp5-Sp9 are structurally similar to other Sp-family members but sequence differences in their *trans*-activation domains consign them to a distinct sub-group.

Sp proteins regulate the promoters of many mammalian genes, including genes controlling cell cycle progression as well as development and tissue differentiation [7,8]. Sp-family members are regulated by a variety of growth-related signal transduction pathways as well as mechanisms controlling embryonic development [9]. Several members of the Sp-family have been shown to encode multiple transcripts and/or protein isoforms. For example, four novel transcripts have been identified in mouse gametes that encode truncated Sp1 proteins [10-13]. *Sp3* encodes three protein isoforms with functional activities that are dependent on the site of translational initiation, whereas Sp4, Sp7 and Sp8 are each encoded by two distinct transcripts [14-20].

Nullizygous or “knock-out” mouse strains are generally employed to determine the physiological function of a particular gene of interest as well as its necessity for development. Although many Sp-family members are expressed broadly during development, results from such “knock-out” mouse strains indicate that the functions of Sp proteins may only partially overlap. For example, *Sp1*-null embryos perish by E10.5-E11. Their gross morphology is abnormal, but many structures and tissues arise without specific developmental defects such as somites, otic vesicle, limb bud, blood, the developing eye, and heart [21]. *Sp3* nullizygotes are 25% smaller than wild-type and heterozygous embryos and expire immediately after birth likely due to cardiac malformations [22,23]. These animals also show incomplete formation of the dentin and enamel layers of the tooth, exhibit hypo-ossification of the head, limb, and trunk bones, and some ossification centers are completely absent [22,24]. Finally, newborn *Sp4* nullizygotes appear normal, albeit smaller than wild-type littermates, yet 70% perish in early post-natal life due to unknown causes [17,25,26].

Sp2 is an enigmatic member of the Sp-family. The Sp2 DNA-binding domain is the least conserved (75%) amongst Sp-family members, Sp2 DNA-binding activity and *trans*-activation are negatively regulated in cultured cells, and the vast majority of Sp2 localizes to sub-nuclear foci associated with the nuclear matrix [6,27]. Using Sp2 protein prepared in insect cells a consensus DNA-binding site (5′-GGGCGGGAC-3′) was identified that is bound with high affinity (K_d = 225 pM) by recombinant Sp2 *in vitro*. Yet, in transient-transfection assays Sp2 only weakly *trans*-activates promoters carrying consensus Sp2-binding sites or well-characterized Sp-dependent promoters that are readily induced by Sp1 and Sp3 [27]. Phan et al. [28] reported that Sp2 abundance is increased significantly in human prostate cancer specimens and the extent of Sp2 expression is directly correlated with the pathological grade of the tumor examined. The zebrafish orthologue of *Sp2* has been cloned and shown to be structurally and functionally analogous to its mammalian counterparts [29]. Zebrafish embryos inherit *Sp2* as a maternal transcript, *Sp2* transcription is triggered in blastula stage embryos, and *Sp2* is required for the completion of gastrulation.

As a first step towards determining if *Sp2* is essential for mouse development we isolated and cloned a mouse *Sp2* cDNA, identified transcriptional start sites, characterized *Sp2* promoters, and identified patterns of *Sp2* expression in developing animals using *in situ* hybridization. Our results indicate that *Sp2* is an alternatively spliced gene and its transcription is governed by at least two promoters that differ in their cell-type specificity. Moreover, a constellation of sub-genomic length messages are synthesized from the *Sp2* locus in a cell- and tissue-specific

fashion and have the capacity to encode a novel, partial Sp2 protein. Finally, full-length *Sp2* message is expressed widely in mid-gestation mouse embryos, particularly in the forebrain, and robust *Sp2* expression continues in neurogenic regions of the post-natal and adult mouse brain.

2. Materials and methods

2.1 Northern blotting

Poly-A(+) mouse multiple-tissue Northern blots (MTN®) were purchased from Clontech, Inc. (Mountain View, CA) and hybridized as recommended by the manufacturer.

2.2 Cloning of mouse Sp2 cDNA

Total RNA was purified from mouse heart (Trizol®, Invitrogen, Inc., Carlsbad, CA), reverse transcribed using oligo-dT primers and SuperScript™ II Reverse Transcriptase (Invitrogen) and subjected to thermal cycling with Titanium™ Taq DNA polymerase (Clontech) and gene-specific primers (5′-CCAGGGAGGAAGATGTCGTAATGAGCGATC-3′ and 5′-GCCAATCAGATGGGAGGTGTTC-3′) complementary to predicted untranslated sequences upstream and downstream of mouse Sp2 coding sequences. A resulting 1911 bp amplified product was sub-cloned in pCR4 using the TOPO® TA Cloning® Kit (Invitrogen) and subjected to automated DNA sequencing. The nucleotide sequence of this cDNA was deposited in GenBank (accession number GU126673).

2.3 5′RACE Assay

The GeneRacer® Kit (Invitrogen) was employed to identify mouse *Sp2* transcriptional start sites precisely as recommended by the manufacturer. First-strand cDNAs were amplified with Titanium™ Taq DNA polymerase (Clontech), a primer complementary to the GeneRacer® RNA oligonucleotide, and nested gene-specific primers complementary to sequences in *Sp2* exons 5 and 7 (5′-GGA CAC CAT TGA TGT TGA TGG TCT GCA-3′ and 5′-CAC GTG CTT CTT CTT GCC CTG C-3′). As a positive control HeLa cell-derived RNA and primers (5′-GCTCACCATGGATGATGATATCGC-3′ and 5′-GACCTGGCCGTCAGGCAGCTCG -3) supplied by the manufacturer were employed to amplify the 5′ end of human β-actin. Amplified products were sub-cloned in pCR4 using the TOPO® TA Cloning® Kit (Invitrogen) and subjected to automated DNA sequencing.

2.4 Cells and cell culture

Human (T98G, DU145, HeLa, HepG2, 293), primate (COS-1), and mouse (F9, L) cell lines were obtained from the ATCC (Manassas, VA). Primary mouse heart fibroblasts and myocytes were a kind gift of Drs. Lianna Li and Barbara Sherry (N.C. State University, Raleigh, NC). Cell lines were cultured in Dulbecco's modified Eagle's medium (DMEM; GIBCO, Inc., Gaithersburg, MD) supplemented with 10% heat-inactivated fetal bovine serum (Atlanta Biologicals, Norcross, GA) and 1% Pipracil at 37°C under 5% CO_2.

2.5 Luciferase constructs

Four genomic fragments (termed Promoter 1 through Promoter 4) derived from the mouse *Sp2* locus were amplified from a 129/SvEv mouse genomic library (Stratagene, Inc., La Jolla, CA) via the PCR using Titanium™ Taq DNA polymerase (Clontech) and the following four pairs of primers: Promoter 1: 5′-CCC GGT ACC GCT AGC CCG AAA CTC ATG ATG GTC C-3′ and 5′-CCC AAG CTT CAC GTC CCC TCA CCC CCT CC-3′; Promoter 2: 5′-GGG GGT ACC GGA AGC TTT GTT ACA TGT TCT-3′ and 5′-GGG GCT AGC CAC CGC CCG CCT AGC ACC CGC CAA T-3′; Promoter 3: 5′-CCC GGT ACC GCA CGG TGC TCT CTG AGG GC-3′ and 5′-CCC AAG CTT TGG TCT GGG AAC TGT TCC CT-3′; Promoter 4: 5′-

CCC GAG CTC CAC ATC TGT AAA CAT GCA TT-3′ and 5′-CCC AAG CTT GTT GAT GTG GTT GCA GCT GT-3′. Amplified products were sub-cloned in pGL3-Basic (Promega Corp., Madison, WI) generating pGL3-P1, pGL3-P2, pGL3-P3 and pGL3-P4 and sequences were confirmed via automated DNA sequencing.

A series of Promoter 2 deletion constructs were prepared via the PCR using circular pGL3-P2 as template, Titanium™ Taq DNA polymerase (Clontech), a single anti-sense 5′ primer (5′-pGGT ACC TAT CGA TAG AGA AA-3′) and the following construct-specific sense primers: pGL3-P2Δ350 (carrying nucleotides from -830 to -1): 5′-pCTA GGA TGC CAG CGT GCA ATT-3′. pGL3-P2 Δ450 (-730 to -1): 5′-pGGA AAT AGC AAA TGT AAC TAG GGA-3′. pGL3-P2Δ565 (-615 to -1): 5′-pCAT AAT GAC AAT TCC CT CTA G-3′. pGL3- P2Δ620 (−560 to -1): 5′-pGAC TCT CGA GAG GCG ATG GAT GG-3′. pGL3-P2Δ678 (-502 to -1): 5′-pGAT GCG CTC ACT GGG GGG CCG-3′. pGL3-P2Δ700 (-480 to -1): 5′-pCGA CAG CTA CAC TTG AGA CTC-3′. pGL3-P2Δ750 (-430 to -1): 5′-pGCT CTT TAC GGG GCA TTT CAC-3′. pGL3-P2Δ778 (-403 to -1): 5′-pGCA CTT CTG ATT GGC TTT AGG G-3′. pGL3-P2Δ817 (-364 to -1): 5′-pGCT GGG TTA GAG GTG GGT GG-3′. pGL3-P2Δ880 (-300 to -1): 5′-pCCT CCT CAC CCA GCT TCC TGT-3′. pGL3-P2Δ1050 (-130 to -1): 5′-pGCT TCC CAA TGA TTG GTT GA-3′. Amplified DNAs were self-ligated and employed in bacterial transformations. The integrity of all luciferase constructs was confirmed by automated DNA sequencing. The nucleotide sequences of Promoters 1 and 2 were deposited in GenBank (accession numbers GU256575 and GU256576).

2.6 Transfection and luciferase assay

One day prior to transfection, 2.0×10^4 cells were plated in each well of 24 well plates. Cells were transfected as described [30]. The Dual-Luciferase® Reporter Assay System (Promega) was employed to quantify luciferase activity precisely as recommended by the manufacturer. Luminescence was detected in a Lumat LB 9507 luminometer (EG&G Berthold), and results were normalized to the abundance of Δ53RL Renilla activity as described [30].

2.7 In vitro protein/DNA-binding assays

Oligonucleotide probes for protein/DNA-binding assays were obtained from Invitrogen, Inc. Oligonucleotides employed include the following and their complementary strands: mSp2-615: 5′-CGC GTA ATG ACA ATT CCC TCT AGG AAG TGC CAG GTA ACC CAA CTC CTG CCT GCC AGG TCG TCC-3′, mSp2-550: 5′-CGC GTG ACT CTC GAG AGG CGA TGG ATG GCG GGC GAG GGC CTG AGG TGA AGG C-3′, mSp2-480: 5′-CGC GTC GAC AGC TAC ACT TGA GAC TCA CCA TGC CAG ACT GTA TTG AGA GAT GTA TAC-3′, mSp2-402: 5′-CGC GTG CAC TTC TGA TTG GCT TTA GGG CAG ATC TAT CAC TTA TCT C-3′. Annealed, complementary oligonucleotides were labeled with [γ-^{32}P] dATP (ICN, Inc., Costa Mesa, CA) and T4 polynucleotide kinase (New England Biolabs, Beverly, MA), and examined for *in vitro* protein/DNA-binding activity as previously described [30].

2.8 Yeast “one-hybrid” screen

The Matchmaker One-Hybrid library Construction & Screening Kit (Clontech Laboratories, Inc., CA) was employed to identify mammalian proteins that bind to portions of *Sp2* Promoter 2. Five DNA fragments were used as “bait” sequences in plasmid pHIS2.1: a 400 bp DNA sequence from -701 to -301, and double-stranded oligonucleotide sequences employed in protein/DNA-binding assays (mSp2-615, mSp2-550, mSp2-480, and mSp2-402). The cDNA library for the “one-hybrid” screen was generated as recommended by the manufacturer from human HeLa cells. Candidate DNAs were analyzed by automated sequencing.

2.9 *In situ* hybridization

Probes for *in situ* hybridization were prepared from a 489 bp portion at the 5′ end of mouse *Sp2* exon 4. This cDNA fragment was isolated via the PCR using Titanium™ Taq DNA polymerase (Clontech) and the following forward and reverse primers: 5′- TAC TAC CCA GGA CTC CCA GCC ATC TC-3′ and 5′-GTA CTT GAC TTC TGG ACA GGA GCT GGC-3′. The amplified fragment was sub-cloned in pSC-A (Stratagene) generating pSC-A-Probe 3. Digoxigenin-labeled probes were prepared using linearized pSC-A-Probe 3, 5× transcription buffer (Roche, Inc., Indianapolis, IN), DIG RNA Labeling Mix (Roche), RNaseIn® (Promega), and RNA polymerase (T7 or Sp6; Roche). Anti-sense or sense probes were synthesized in these reactions depending on the orientation of the cDNA insert within pSC-A. Labeled probes were purified from unincorporated nucleotides with NucAway™ Spin Columns (Ambion) according to the manufacturer's instructions. *In situ* hybridization of tissue sections was performed essentially as described [31].

3. Results

3.1 Cloning and characterization of mouse Sp2 cDNAs

The sequences of ten mouse *Sp2* cDNAs have been deposited in the GenBank database. Alignments of these cDNAs with mouse genomic DNA revealed that each shares an identical structure beginning with the presumed first translated exon (termed exon 3) through the final coding exon (termed exon 8). However, these cDNAs form three distinct groups based on sequences carried upstream of exon 3. The longest *Sp2* cDNA, isolated from a mouse testis library (AK019649), features an extended 5′ untranslated region that is derived from two exons, termed exons 1 and 2, that are approximately 20 kbp upstream of exon 3 (Fig. 1A). Two additional groups of *Sp2* cDNAs lack exons 1 and 2 and carry single novel exons (termed exon 2A or 2B) that are 14 kbp and 5 kbp upstream of exon 3, respectively (Fig. 1A). Exon 2A encodes an in-frame translational start site that if utilized is predicted to add five amino acids to the amino-terminus of Sp2 protein. The structural differences carried by these three groups of mouse *Sp2* cDNAs indicate that the 5′ end of the *Sp2* locus is subject to alternative splicing.

To begin a detailed characterization of *Sp2* expression, a mouse *Sp2* cDNA was amplified from total heart RNA utilizing the PCR and gene-specific primers that bound sequences within exon 2A (mSp2-5′) and exon 8 (mSp2-3′). A single amplification product was produced and this cDNA, termed mSp2, was sequenced in its entirety and deposited in GenBank (GU126673). The mSp2 cDNA is 1911 bp in length, and carries an open reading frame predicted to encode a protein of 612 amino acids. Comparison of the sequence of this cDNA with sequences obtained previously from specific (AK078246, olfactory region; AK139302, cerebellum; AK156580, activated spleen; AK080877 neonatal adipose tissue; AK019649, testis) or mixed-tissue (AK172877; BC021759; BC086457; NM_001080964; NM_030220) mouse libraries identified only five single nucleotide polymorphisms within the coding regions of exons 3-8. The coding region of the mSp2 heart-derived cDNA is identical to that of cDNA AK156580 and carry two single nucleotide polymorphisms that are not shared by the other nine mouse *Sp2* cDNAs analyzed: (1) a guanine residue at nucleotide 155 (with the adenine within the translational start site in exon 3 designated as +1) instead of a cytosine, resulting in the conversion of an alanine residue within the *trans*-activation domain to glycine. (2) A thymidine at nucleotide 497 instead of a cytosine, converting a threonine residue within the *trans*-activation domain to isoleucine.

In contrast to the near uniformity of mouse cDNA sequences within *Sp2* exons 3-8, cDNA sequences upstream of exon 3 are considerably more diverse. Five mouse cDNAs (AK078246, AK156580, AK139302, AK019649, and NM_030220) include three additional nucleotides upstream of exon 3 presumably due to the utilization of a closely-juxtaposed, distinct splice

acceptor site. Translation of these three nucleotides is predicted to result in the insertion of an alanine residue into the mouse *Sp2* coding sequence if translation begins at a start site in exon 2A that is carried by each of these cDNAs. Two additional *Sp2* cDNAs (AK080877 and AK019649) carry sequences upstream of exon 3 that are completely distinct from other mouse *Sp2* cDNAs analyzed, suggesting that these sequences result from alternative splicing. As will be discussed shortly, this suggestion is supported by comparisons with sequences upstream of exon 3 obtained from 5′RACE studies.

3.2 A constellation of full-length and partial transcripts are synthesized from the mouse Sp2 locus

As mentioned above, *Sp2* cDNAs recovered by RT-PCR and from various mouse libraries indicate that alternative splicing at the 5′ end of the *Sp2* locus leads to the synthesis of three groups of transcripts. To determine if additional *Sp2* locus-derived transcripts are synthesized *in vivo*, a series of Northern blotting experiments were performed using radiolabeled probes and poly-A (+) RNAs prepared from nine adult mouse organs and whole embryos. Two non-overlapping probes were employed that were derived from portions of the mSp2 *trans*-activation domain that share less than 50% sequence homology with other members of the Sp-family. Probe 1 includes 19 bases from the 3′ end of exon 3 as well as the first 470 bases of exon 4 (exon 4 is 963 bases in length). Probe 2 includes the 3′ 174 bases of exon 4, all of exon 5 sequences, as well as 16 nucleotides from the 5′ end of exon 6. Both radiolabeled probes detected a 4.7 kb transcript in all RNA samples analyzed, however probe 1 hybridized to this transcript more efficiently than probe 2 (arrows, Fig. 1B, C). The abundance of this transcript varied across the samples analyzed, the highest levels of expression were noted in embryos and lowest levels were detected in kidney-derived RNAs. In addition to this 4.7 kb transcript, each probe detected a number of smaller transcripts ranging in size from 2.7 to 1.5 kb (brackets, Fig. 1B, C). The abundance and sizes of these transcripts varied across the samples analyzed, with the highest levels of expression noted in RNAs derived from thymus, brain, and liver. Interestingly, despite its relatively inefficient hybridization to the 4.7 kb *Sp2* transcript probe 2 bound these smaller *Sp2* locus-derived transcripts as efficiently as probe 1 (Fig. 1C). Given the absence of evidence for alternative splicing downstream of exon 3, these data suggested that the *Sp2* locus may encode multiple promoters and/or generate messages of distinct sizes via one or more alternative mechanisms.

3.3 Transcriptional start sites within the mouse Sp2 locus are utilized in a tissue-specific fashion

To determine if the constellation of *Sp2* transcripts detected in Northern blots have distinct 5′ ends, a series of 5′RACE experiments were performed using nested gene-specific primers derived from exons 5 and 7. Total RNAs were prepared from mouse heart, liver, lung, epidermis, spleen, kidney, and brain, and each RNA sample was analyzed by 5′RACE. As shown in Fig. 2A, RNAs from each mouse tissue analyzed produced a unique pattern of 5′ RACE products. A majority of the 5′RACE products produced from mouse heart, liver, and spleen RNAs were cloned and sequenced to confirm that they were derived from the *Sp2* locus and to define their 5′ ends. DNA sequencing confirmed that each 5′RACE product was derived from the *Sp2* locus and revealed a bevy of 5′ ends that clustered within three distinct regions. The first cluster consists of 5′ RACE products of approximately 1380 bp that were detected in reactions utilizing total RNAs from mouse heart and liver (indicated by a filled arrowhead, Fig. 2A). Sequencing of these 5′RACE products revealed that their 5′ ends are similar to previously isolated mouse *Sp2* cDNAs that carry exon 2A, including the aforementioned heart-derived mSp2 cDNA isolated by RT-PCR as well as several sequences deposited in GenBank (AK172877, BC021759, NM_001080964, and NM_030220). We defined the 5′ end of exon 2A as being the first nucleotide of the longest 5′RACE product (obtained from mouse liver), and noted that heart-derived 5′RACE products initiated 28 and 61 bp downstream. Transcripts

initiating within this 61 bp region of exon 2A were termed type I transcripts, and are predicted to encode full-length Sp2 protein (Fig. 2C).

A second cluster of 5′ ends, representing mRNAs termed type II transcripts, initiated within exon 4 (Fig. 2C). RNAs from mouse heart, liver, and spleen produced unique patterns of 5′ RACE products that initiated within a broad 615 bp region of exon 4. Mouse heart tissue produced unique 5′RACE products of approximately 895 bp (open arrowhead, Fig. 2A) that initiate at one of two nucleotide positions 340 bp from the 5′ end of exon 4. RNAs prepared from mouse liver produced a bevy of type II 5′RACE products that initiate 620 bp, 830 bp, and 920 bp from the 5′ end of exon 4, as well as a product that initiates 953 bp from the 5′ end of exon 4 that was also amongst 5′RACE products recovered from heart and spleen. Each of these type II transcripts initiates within an exon that encodes the vast majority of the *Sp2 trans*-activation domain and thus hypothetical proteins synthesized from such transcripts would be expected to be functionally distinct from full-length Sp2. In this vein, the first potential in-frame translational initiation site (5′-cag-gcc-ATG-cag-3′) in such transcripts is not encountered until the distal end of exon 5 (Fig. 2C). Should this translational start site be utilized it is predicted to result in the synthesis of a 178 amino acid partial-Sp2 protein, with a predicted molecular weight of 20 kDa, that carries a small portion of the *trans*-activation domain and the entirety of the DNA-binding domain.

A final cluster of 5′RACE products, termed type III transcripts, were recovered from mouse liver and initiated at two positions (83 bp and 144 bp from the 5′ end of exon 5; Fig. 2C). A similar 5′RACE product has also been obtained from mouse brain tissue and initiated 87 bp from the 5′ end of exon 5 (data not shown). As for type II transcripts, an in-frame translational start site within exon 5 is carried by type III transcripts and is predicted to encode a 20 kDa partial-Sp2 protein.

3.4 Transcriptional start sites within the mouse Sp2 locus are utilized in a cell type-specific fashion

To determine if *Sp2* transcripts may also be synthesized in a cell type-specific fashion, primary cultures of purified mouse cardiac myocytes or fibroblasts were obtained. The cardiac myocyte cultures utilized have been shown to be greater than 95% pure, whereas the cardiac fibroblast cultures have been shown to be more than 99% pure [32]. Total RNAs were prepared from each culture and subjected to 5′RACE analysis, and resulting products were cloned and sequenced. Consistent with the notion that patterns of *Sp2* transcription are indeed cell type-specific, type I and type III transcripts were detected in RNAs prepared from cardiac fibroblasts whereas only type II transcripts were detected in cardiac myocytes. As shown in Figs. 2B and C, a type I 5′RACE product of 1350 bp (filled arrowhead) was detected in mouse cardiac fibroblasts that initiates 61 bp from the 5′ end of exon 2A. Additionally, a unique 5′RACE product recovered from cardiac fibroblast cultures carried 5′ untranslated sequences upstream of exon 3 that are nearly identical to a novel mouse *Sp2* cDNA sequence deposited in GenBank (AK080877; Fig. 1A). These sequences (termed exon 2B) replace nucleotides encoded by exon 2A, resulting in the elimination of the putative translational start site at the 3′ end of this exon. These results confirm that the 5′ end of the *Sp2* locus is subject to alternative splicing and suggest that *Sp2* is spliced in a cell type-specific fashion. Type III 5′RACE products recovered from cardiac fibroblast cultures initiated 86 bp from the 5′ end of exon 5 (Fig. 2C). In contrast to results for cardiac fibroblast cultures, only type II 5′RACE products were recovered from cardiac myocyte cultures and these messages initiated 397 bp and 837 bp from the 5′ end of exon 4 (Fig. 2C).

We conclude from 5′RACE studies that transcriptional initiation within the mouse *Sp2* locus is both tissue- and cell type-specific and produces more than three distinct classes of transcripts. Transcriptional start sites within exon 2A produce type I transcripts and encode full-length Sp2

protein via translational initiation within exons 2A and/or 3. Exon 2B-containing transcripts would be expected to produce full-length Sp2 protein as well. Transcriptional initiation at a plethora of sites within exons 4 and 5 leads to the production of type II and type III transcripts, respectively, and these mRNAs have the potential to encode a partial Sp2 protein.

3.5 Mouse Sp2 transcription is regulated by at least two promoters

Given 5′RACE results and the aforementioned structural differences at the 5′ ends of mouse *Sp2* cDNAs, we reasoned that the *Sp2* locus may be regulated by multiple promoters. To test this hypothesis genomic DNA fragments preceding transcriptional initiation sites for type I (Promoter 2; pGL3-P2), II (Promoter 3; pGL3-P3), and III (Promoter 4; pGL3-P4) transcripts were subcloned into pGL3-Basic, a vector that includes the gene for firefly luciferase and lacks eukaryotic promoter or enhancer elements. Genomic sequences upstream of *Sp2* exon 1 (Promoter 1; pGL3-P1) were also subcloned into pGL3-Basic to determine if promoter activity could be detected. Each of these reporter constructs was employed in a series of transient transfection experiments that utilized six mammalian cell lines (293, COS-1, DU145, HeLa, T98G, F9) as well as primary cell cultures (mouse heart-derived fibroblasts and myocytes). Resulting luciferase activities were normalized to a *Renilla* luciferase reporter gene as a control for plate-to-plate variations in transfection efficiency. As illustrated in Fig. 3, the pGL3-Basic construct carrying Promoter 2 sequences stimulated luciferase expression 28- to 183-fold in the cell lines examined and 62-fold in primary heart-derived cell cultures. With the exception of COS-1 cells, cultures receiving a pGL3-Basic reporter gene carrying Promoter 1-derived genomic sequences showed little or no luciferase activity. Luciferase expression directed by Promoter 1 sequences in COS-1 cells was quite significant (22-fold increase in relative *trans*-activation), achieving nearly 40% of the level of *trans*-activation directed by Promoter 2 sequences in this cell line. In stark contrast to results with Promoters 1 and 2, little or no firefly luciferase activity was detected in cultures that received pGL3-Basic reporter genes prepared with Promoter 3 or 4 genomic sequences. We conclude from these transient transfection results that the Promoter 2 genomic fragment directs transcription robustly in a variety of primate and mouse cell lines as well as primary cell cultures derived from mouse heart. Other *Sp2* locus-derived genomic sequences direct transcription in a highly circumscribed, cell type-specific fashion (Promoter 1) or exhibited little or no activity in parallel assays (Promoters 3 and 4).

To delineate promoter elements that govern Promoter 2 activity, a series of nested 5′-deletions within pGL3-P2 were generated by the PCR, and luciferase activities were quantified relative to that of *Renilla* luciferase in mammalian cell lines and strains. The transcriptional activities shown in Fig. 4 are derived from the average activity of each deletion construct in all cell types relative to the activity of pGL3-P2 (set equal to 100). The activity of pGL3-P2Δ350, in which 350 nucleotides were deleted from the distal end of Promoter 2, was nearly identical to that of pGL3-P2. Deletion of an additional 100 nucleotides reduced luciferase activity (to 65% relative to pGL3-P2), and deletion of a further 75 nucleotides marginally increased luciferase activity (to 89% relative to pGL3-P2). In contrast to these relatively modest changes in *trans*-activation, pGL3-P2Δ620 transcriptional activity was reduced markedly (to 3% relative to pGL3-P2). These results indicate that one or more elements between -615 and -560 play an important role in stimulating Promoter 2 transcriptional activity. Deletion of a further 58 nucleotides (pGL3-P2Δ678) resulted in a marked stimulation of Promoter 2 activity (to 118% relative to pGL3-P2), implying that sequences between -560 and -502 function in conjunction with cognate DNA-binding proteins as a transcriptional repressor. A further deletion of 22 nucleotides (pGL3-P2 Δ700) nearly doubled levels of *trans*-activation (190% relative to pGL3-P2). Deletion of an additional 50 nucleotides (pGL3-P2Δ750) virtually eliminated luciferase activity (1% relative to pGL3-P2) and a further deletion of 28 nucleotides produced a construct (pGL3-P2Δ778) with the lowest levels of transcriptional activity of all reporter constructs examined (only 0.1% of the activity of pGL3-P2). These results indicate that one or more

elements between -480 and -430 stimulate Promoter 2 transcriptional activity. Elimination of an additional 39 nucleotides (pGL3-P2Δ817) returned transcriptional activity to levels akin to that of wild-type Promoter 2 (116% relative to pGL3-P2), thus additional inhibitory sequences were mapped between -402 and -363. The final two deletion constructs, pGL3-P2Δ880 and pGL3-P2Δ1050, exhibited relative transcriptional activities that were lower than that of pGL3-P2 (63 and 39%). These results indicate that the region deleted between pGL3-P2Δ817 and pGL3-P2Δ1050 encodes one or more relatively weak elements that stimulate basal transcription activity. Based on the transcriptional activities of these nested deletions, five discrete promoter sequences that govern Sp2 Promoter 2 activity were identified. Three (-830 to -560, -502 to -430, and -363 to -130) stimulate Promoter 2 transcriptional activity and two (-560 to -502 and -430 to -363) repress Promoter 2 transcriptional activity.

With the delineation of five functionally important portions of Promoter 2, four approaches were undertaken to identify *trans*-acting factors that may regulate their activities. First, four double-stranded oligonucleotides derived from putative stimulatory (mSp2-615, mSp2-480) or repressive (mSp2-550, mSp2-402) Promoter 2 elements were synthesized, radiolabeled, and employed in *in vitro* protein/DNA-binding assays with nuclear or whole-cell extracts prepared from DU145 and COS-1 cells. Such assays failed to identify protein/DNA complexes that bound these sequences specifically under a variety of experimental conditions (data not shown). Second, each Promoter 2-derived element or a 400 bp Promoter 2 fragment that carries all five elements were employed as “baits” in a series of yeast “one-hybrid” assays in conjunction with a cDNA library prepared from HeLa cells. These assays, once again, failed to identify *trans*-acting factors that bound specifically to one or more portions of Promoter 2 (data not shown). Third, putative transcription factor-binding sites within the 1180 bp Promoter 2 genomic fragment were identified via computational analysis (MatInspector Release professional 7.7.3.1 with matrix library 8.0). This *in silico* analysis identified a number of putative transcription factor-binding sites, some of which localized to regions shown to be of functional importance. For example, the proximal region (from -200 to +2) of the promoter was predicted to carry four CCAAT or CCAAT/NF-Y boxes (at +2, -41, -83, -121) and GC-rich sequences that may be targeted by Sp-family members (at +21 and -15; Fig. 4). Further upstream, a predicted Sp2-binding site was identified at -225, an LTR-derived CCAAT box was identified at -298, and a predicted site for the binding of CDP/Cut factors was noted at -366 (Fig. 4). A canonical TATA box was not identified in this analysis. Finally, the mouse Promoter 2 sequence was aligned with analogous genomic sequences derived from 15 mammals, including the dog, horse, pika, opossum, megabet, microbet, orangutan, dolphin, cow, macaque, hyrax, guinea pig, rat, chimpanzee, and human. This alignment revealed that the proximal 400 bp of the promoter region exhibited 66-80% identity (data not shown), and nucleotides within a subset of putative transcription factor-binding sites (*i.e.*, GC-rich sequences at +21, CCAAT/NF-Y boxes at +2 and -41, Sp2-binding site at -225, LTR-derived CCAAT box at -298, and predicted CDP/Cut-binding site at -366) are identical in all species examined (data not shown and Fig. 4). Additionally, GC-rich sequences at -15, a CCAAT box at -83, and a predicted CCAAT/NF-Y box at -121 were shared by 11 to 13 of the mammalian species examined (data not shown and Fig. 4). We conclude from these studies that Promoter 2 carries at least five functionally significant elements, three stimulatory and two repressive, that contribute to *in vitro trans*-activation and these elements overlap with a handful of predicted transcription factor-binding sites. Additional studies will be required to determine whether one or more of these *trans*-acting factors regulates *Sp2* transcription.

3.6 Detection of Sp2 expression via *in situ* hybridization

It is well established that *Sp2* is expressed in many, if not all, mammalian cell lines and the evidence presented thus far indicates that *Sp2* message is expressed in many mouse organs. Yet, these results do not address whether *Sp2* is expressed in all or only a fraction of cells

within a given organ or tissue nor do they speak to whether *Sp2* is expressed in a developmentally-regulated fashion. To address these issues, *in situ* hybridization was performed on embryonic and adult mouse tissues. Sense and anti-sense digoxigenin-labeled *Sp2* probes utilized for these studies were derived from a 489 bp portion of the 5′ end of exon 4 (Probe 3) that is predicted to detect transcripts that encode full-length Sp2 protein (*i.e.*, transcripts that include exon 3) as well as transcripts initiating at the extreme 5′ end of exon 4 (Fig. 2C). Type II transcripts initiating within the 3′ half of exon 4 or type III transcripts were predicted to be excluded from this analysis.

Mouse embryos at day 13 (E13) of gestation were collected, fixed, and sagittal sections were prepared for hybridization. As shown in Fig. 5A, an anti-sense probe detected *Sp2* expression throughout E13 embryos with the exception of the developing mouse skeleton (filled arrowhead). *Sp2* expression was most abundant in the embryonic brain, particularly in the forebrain (Fig. 5A, open arrowhead). In contrast to these findings, a sense probe did not stain mouse embryo sections examined in parallel (Fig. 5B). It is worth noting that a virtually identical distribution of *Sp2* expression has been detected in developing zebrafish embryos using a digoxigenin-labeled RNA probe derived from the zebrafish *Sp2* 3′ untranslated region [29].

Given the relatively high levels of *Sp2* expression in the embryonic brain it became of interest to determine whether *Sp2* is expressed in the brains of post-natal animals. Post-natal day 5 animals were perfused, fixed with paraformaldehyde, and mouse brains were collected, frozen, cut into sagittal sections, and hybridized with sense and anti-sense digoxigenin-labeled RNAs. Microscopic examination of brain sections stained with an *Sp2* anti-sense probe at low power revealed strong hybridization at sites where post-natal neurogenesis is persistent (Fig. 6A). One site was the cerebellum, where the external (EGL) and internal granule (IGL) layers were highly stained whereas the Purkinje cell layer (PCL) was relatively poorly stained with this probe (Fig. 6C). Additionally, robust *Sp2* expression was detected at sites that support persistent neurogenesis in the olfactory bulb (OLB), the subventricular zone (SVZ) of the lateral ventricles, the rostral-migratory stream (RMS), as well as the dentate gyrus (DG) of the hippocampus (Figs. 6A and C). Consistent with results from embryonic sections, an *Sp2* sense probe did not hybridize to these or other regions of the mouse brain (Fig. 6B).

Given that *Sp2* expression in post-natal animals appeared to be limited to regions engaged in neurogenesis it became of interest to determine whether *Sp2* expression continues in these regions in adult animals. *In situ* hybridization of the adult mouse brain revealed that *Sp2* is expressed strongly in the DG and CA1 regions of the hippocampus, as well as the granule cell (GCL), mitral cell (MCL), and glomerular (GL) layers of the OLB, the SVZ and RMS, and the IGL of the cerebellum (Fig. 7A and C, and data not shown). Relatively weak levels of *Sp2* expression were detected within the CA3 region of the hippocampus in the adult mouse brain (Fig. 7C, center panel). In contrast to *in situ* hybridization results from post-natal animals, *Sp2* expression was not detected within the molecular layer (ML) of the cerebellum that is formed by trailing axons from the IGL [33-35]. The continued expression of *Sp2* in progenitor regions of the adult brain (*e.g.*, dentate gyrus of the hippocampus, as well as the SVZ and RMS), suggest a potential role for this transcription factor in neuronal cell proliferation and/or migration. Expression of *Sp2* in maturing regions of these structures (*e.g.*, internal granule layer of the cerebellum, olfactory bulb layers, and the CA fields of the hippocampus) may also indicate that *Sp2* functions during later stages of development (*e.g.*, process arborization and remodeling) that are unique to these structures.

4. Discussion

Sp-family members have been implicated in the transcriptional regulation of genes required for development, cell-cycle progression, and tumorigenesis. Yet, molecular and biochemical studies focusing on Sp2 function have yielded few insights into key target genes nor its roles in cell and/or organismal physiology. To begin to address these issues and as a prelude to the generation of nullizygous *Sp2* animals, we performed a detailed examination of mouse *Sp2* transcription and report that (1) the 5′ untranslated region of *Sp2* is subject to alternative splicing, (2) *Sp2* transcription is regulated by at least two promoters that differ in their cell-type specificity, (3) an *Sp2* promoter that is active in nine mammalian cell lines and cell strains is regulated by at least five discrete enhancer and inhibitory elements, (4) sub-genomic messages that have the capacity to encode a novel partial-Sp2 protein are synthesized from the *Sp2* locus in a tissue- and cell type-specific fashion, and (5) *Sp2* is expressed widely during embryogenesis, particularly in the embryonic brain, and robust *Sp2* expression occurs in neurogenic regions of the post-natal and adult mouse brain.

Two classes of alternatively spliced *Sp2* cDNAs were isolated and characterized in this study and a testis-specific, alternatively-spliced transcript (AK019649) has been identified previously. Alternative splicing of these cDNAs affects the 5′ untranslated region of *Sp2* linking alternative upstream exons to the presumed first coding exon (exon 3). Transcripts analogous to AK019649 were not detected in this study using RNAs derived from more than a half dozen mouse tissues as well as cultured primary cells. Consistent with the circumscribed expression of AK019649, a 1.6 kbp genomic fragment (Promoter 1) immediately upstream of exon 1 exhibited significant (22-fold increase in relative *trans*-activation) promoter activity in only one (COS-1) of nine mammalian cell lines and cell strains examined. In contrast, *Sp2* cDNAs carrying exon 2A (type I transcripts) have been recovered previously from a wider range of mouse tissues and were recovered from mouse heart, kidney, and liver in this study via 5′RACE. With that said, type I transcripts were detected by 5′RACE in only half of the adult mouse tissues examined and their synthesis was cell type-specific (*i.e.*, heart-derived fibroblast cultures but not heart-derived myocyte cultures). These results suggest that type I transcripts may be quite limited in some adult mouse tissues. A third alternatively-spliced *Sp2* message that carries exon 2B was isolated from heart-derived fibroblasts and a closely-related transcript (AK080877) has been recovered previously from a mouse neonatal adipose tissue cDNA library. Exon 2B is located 9 kbp downstream from Promoter 2 and thus its transcription is likely to be directed via a distinct, and as yet uncharacterized, promoter. Is alternative splicing of exons encoding the 5′ untranslated region of *Sp2* of functional significance? Translation of exon 2A-containing messages has the potential to append five novel amino acids to the Sp2 amino-terminus due to utilization of an in-frame translational initiation site located at the 3′ end of this exon. Should this initiation site be utilized *in vivo* it is at least conceivable that these additional amino acids could affect Sp2 function, localization, and/or stability. It is also possible that alternative splicing of the 5′ untranslated region generates *Sp2* messages that differ in their stability and/or efficiency of translation. Additional experiments will be required to assess each of these possibilities.

In addition to complexities introduced by alternative splicing, our studies indicate that sub-genomic messages (type II and III transcripts) are initiated within specific *Sp2* exons and the synthesis of these messages is tissue- and cell-specific. It has become increasingly apparent that the synthesis of such messages is quite common in the mammalian genome. In a genome-wide analysis of transcriptional start sites 34-48% of human and mouse genes, including *Sp2*, were reported to carry alternative promoters that overlap with coding sequences [36]. Moreover, such transcriptional start sites were found to cluster and could be classified into groups based on the distribution of transcriptional start sites within a cluster. Our results indicate that the mouse *Sp2* locus harbors at least two clusters of transcriptional start sites that

overlap exons 4 and 5, and these clustered transcripts are distributed over a relatively broad range of nucleotides (*e.g.*, 600 bp for type II transcripts). Is there reason to suspect that sub-genomic Sp2 transcripts may be biologically significant? Translation beginning at a single in-frame translational initiation site carried by type II and III transcripts is predicted to produce a partial-Sp2 protein that, akin to Sp3 isoforms generated by internal translational initiation, carries a portion of the *trans*-activation domain and the entirety of the DNA-binding domain. As for Sp3 isoforms, the synthesis of this partial–Sp2 protein would be predicted to function as a dominant-negative and thus have significant consequences for the regulation of Sp2 target genes. Unfortunately, this predicted partial-Sp2 protein has not as yet been detected due to the absence of appropriate antisera. It is also conceivable that sub-genomic *Sp2* messages are not translated and have functions akin to other non-coding RNAs that have been described [37-41].

A 2.0 kbp genomic fragment (Promoter 2) immediately proximal to exon 2A exhibited robust (28-183-fold increase in relative *trans*-activation) promoter activity in all nine cell lines and cell strains examined. Nested deletions within Promoter 2 indicate that transcription directed by this sequence, at least *in vitro*, is governed by five discrete stimulatory and inhibitory elements arrayed in an alternating pattern. These elements exhibited stimulatory and repressing activities in all nine cell lines and cell strains examined, and thus the factors that regulate these elements may be expressed broadly. Yet, these transient transfection results also indicate that *Sp2* transcription may be regulated by additional regulatory elements. For example, Promoter 2 and its derivatives exhibited identical transcriptional activities in heart-derived fibroblasts and myocytes yet transcripts carrying exon 2A were not detected in myocytes via 5′RACE. This apparent contradiction suggests that additional genomic sequences regulating cell-specific Promoter 2 activity may remain to be discovered. Similarly, this deficiency may also explain why genomic fragments isolated from regions upstream of type II and type III transcripts exhibited little or no transcriptional activity *in vitro*.

Regardless of the potential functionality of sub—genomic *Sp2* transcripts, our results indicate that the expression of full-length *Sp2* message is widespread in mid-gestation mouse embryos and thus is likely to play a role in mouse development. In keeping with this supposition we have cloned the zebrafish *Sp2* orthologue and shown that *Sp2* is required for the completion of gastrulation [29]. Expression of full-length *Sp2* message is particularly prominent in the embryonic forebrain of the mouse, suggesting that Sp2 may regulate region-specific genes within the developing central nervous system. Consistent with a potential role for Sp2 in regional brain development and maturation, robust *Sp2* expression was detected in neurogenic regions of the post-natal and adult mouse brain. Detailed analyses of nullizygous animals will be required to determine if Sp2 function is essential for mouse development and/or organogenesis and to identify key target genes.

Acknowledgments

The authors thank members of the Horowitz laboratory for helpful discussions. This work was funded in part by National Institutes of Health grants GM065405 and CA105313 (to J.M.H.) and funds supplied by the Jimmy V-NCSU Cancer Therapeutics Training Program.

References

1. Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, Dewar K, Doyle M, FitzHugh W, Funke R, Gage D, Harris K, Heaford A, Howland J, Kann L, Lehoczky J, LeVine R, McEwan P, McKernan K, Meldrim J, Mesirov JP, Miranda C, Morris W, Naylor J, Raymond C, Rosetti M, Santos R, Sheridan A, Sougnez C, Stange-Thomann N, Stojanovic N, Subramanian A, Wyman D, Rogers J, Sulston J, Ainscough R, Beck S, Bentley D, Burton J, Clee C, Carter N, Coulson A, Deadman R, Deloukas P, Dunham A, Dunham I, Durbin R, French L, Grafham D, Gregory S, Hubbard T,

Humphray S, Hunt A, Jones M, Lloyd C, McMurray A, Matthews L, Mercer S, Milne S, Mullikin JC, Mungall A, Plumb R, Ross M, Shownkeen R, Sims S, Waterston RH, Wilson RK, Hillier LW, McPherson JD, Marra MA, Mardis ER, Fulton LA, Chinwalla AT, Pepin KH, Gish WR, Chissoe SL, Wendl MC, Delehaunty KD, Miner TL, Delehaunty A, Kramer JB, Cook LL, Fulton RS, Johnson DL, Minx PJ, Clifton SW, Hawkins T, Branscomb E, Predki P, Richardson P, Wenning S, Slezak T, Doggett N, Cheng JF, Olsen A, Lucas S, Elkin C, Uberbacher E, Frazier M, Gibbs RA, Muzny DM, Scherer SE, Bouck JB, Sodergren EJ, Worley KC, Rives CM, Gorrell JH, Metzker ML, Naylor SL, Kucherlapati RS, Nelson DL, Weinstock GM, Sakaki Y, Fujiyama A, Hattori M, Yada T, Toyoda A, Itoh T, Kawagoe C, Watanabe H, Totoki Y, Taylor T, Weissenbach J, Heilig R, Saurin W, Artiguenave F, Brottier P, Bruls T, Pelletier E, Robert C, Wincker P, Smith DR, Doucette-Stamm L, Rubenfield M, Weinstock K, Lee HM, Dubois J, Rosenthal A, Platzer M, Nyakatura G, Taudien S, Rump A, Yang H, Yu J, Wang J, Huang G, Gu J, Hood L, Rowen L, Madan A, Qin S, Davis RW, Federspiel NA, Abola AP, Proctor MJ, Myers RM, Schmutz J, Dickson M, Grimwood J, Cox DR, Olson MV, Kaul R, Raymond C, Shimizu N, Kawasaki K, Minoshima S, Evans GA, Athanasiou M, Schultz R, Roe BA, Chen F, Pan H, Ramser J, Lehrach H, Reinhardt R, McCombie WR, de la BM, Dedhia N, Blocker H, Hornischer K, Nordsiek G, Agarwala R, Aravind L, Bailey JA, Bateman A, Batzoglou S, Birney E, Bork P, Brown DG, Burge CB, Cerutti L, Chen HC, Church D, Clamp M, Copley RR, Doerks T, Eddy SR, Eichler EE, Furey TS, Galagan J, Gilbert JG, Harmon C, Hayashizaki Y, Haussler D, Hermjakob H, Hokamp K, Jang W, Johnson LS, Jones TA, Kasif S, Kaspryzk A, Kennedy S, Kent WJ, Kitts P, Koonin EV, Korf I, Kulp D, Lancet D, Lowe TM, McLysaght A, Mikkelsen T, Moran JV, Mulder N, Pollara VJ, Ponting CP, Schuler G, Schultz J, Slater G, Smit AF, Stupka E, Szustakowski J, Thierry-Mieg D, Thierry-Mieg J, Wagner L, Wallis J, Wheeler R, Williams A, Wolf YI, Wolfe KH, Yang SP, Yeh RF, Collins F, Guyer MS, Peterson J, Felsenfeld A, Wetterstrand KA, Patrinos A, Morgan MJ, de JP, Catanese JJ, Osoegawa K, Shizuya H, Choi S, Chen YJ. Initial sequencing and analysis of the human genome. Nature 2001;409:860–921. [PubMed: 11237011]

2. Tupler R, Perini G, Green MR. Expressing the human genome. Nature 2001;409:832–833. [PubMed: 11237001]
3. Kadonaga JT, Courey AJ, Ladika J, Tjian R. Distinct regions of Sp1 modulate DNA binding and transcriptional activation. Science 1988;242:1566–1570. [PubMed: 3059495]
4. Courey AJ, Tjian R. Analysis of Sp-1 in vivo reveals multiple transcriptional domains, including a novel glutamine-rich activation motif. Cell 1988;55:887–898. [PubMed: 3142690]
5. Kadonaga JT, Carner KR, Masiarz FR, Tjian R. Isolation of cDNA encoding transcription factor Sp1 and functional analysis of the DNA binding domain. Cell 1987;51:1079–1090. [PubMed: 3319186]
6. Moorefield KS, Yin H, Nichols TD, Cathcart C, Simmons SO, Horowitz JM. Sp2 localizes to sub-nuclear foci associated with the nuclear matrix. Mol Biol Cell 2006;17:1711–1722. [PubMed: 16467376]
7. Bouwman P, Philipsen S. Regulation of the activity of Sp1-related transcription factors, Mol. Cell Endocrinol 2002;195:27–38.
8. Safe S, Abdelrahim M. Sp transcription factor family and its role in cancer. Eur J Cancer 2005;41:2438–2448. [PubMed: 16209919]
9. Zhao C, Meng A. Sp1-like transcription factors are regulators of embryonic development in vertebrates. Develop Growth Differ 2005;47:201–211.
10. Persengiev SP, Saffer JD, Kilpatrick DL. An alternatively spliced form of the transcription factor Sp1 containing only a single glutamine-rich transactivation domain. Proc Natl Acad Sci USA 1995;92:9107–9111. [PubMed: 7568082]
11. Persengiev SP, Raval PJ, Rabinovitch S, Millette CF, Kilpatrick DL. Transcription factor Sp1 is expressed by three different developmentally regulated messenger ribonucleic acids in mouse spermatogenic cells. Endocrinology 1996;137:638–646. [PubMed: 8593813]
12. Thomas K, Wu J, Sung DY, Thompson W, Powell M, McCarrey J, Gibbs R, Walker W. SP1 transcription factors in male germ cell development and differentiation. Mol Cell Endocrinol 2007;270:1–7. [PubMed: 17462816]
13. Thomas K, Sung DY, Yang J, Johnson K, Thompson W, Millette C, McCarrey J, Breitberg A, Gibbs R, Walker W. Identification, characterization, and functional analysis of sp1 transcript variants expressed in germ cells during mouse spermatogenesis. Biol Reprod 2005;72:898–907. [PubMed: 15601926]

14. Kennett SB, Udvadia AJ, Horowitz JM. Sp3 encodes multiple proteins that differ in their capacity to stimulate or repress transcription. Nucl Acids Res 1997;25:3110–3117. [PubMed: 9224612]
15. Kennett SB, Moorefield KS, Horowitz JM. Sp3 represses gene expression via the titration of promoter-specific transcription factors. J Biol Chem 2002;277:9780–9789. [PubMed: 11773047]
16. Spengler ML, Kennett SB, Moorefield KS, Simmons SO, Brattain MG, Horowitz JM. Sumoylation of internally initiated Sp3 isoforms regulates transcriptional repression via a Trichostatin A-insensitive mechanism. Cell Signal 2005;17:153–166. [PubMed: 15494207]
17. Supp DM, Witte DP, Branford WW, Smith EP, Potter SS. Sp4, a member of the Sp1-family of zinc finger transcription factors, is required for normal murine growth, viability, and male fertility. Devel Biol 1996;176:284–299. [PubMed: 8660867]
18. Milona MA, Gough JE, Edgar AJ. Expression of alternatively spliced isoforms of human Sp7 in osteoblast-like cells. BMC Genomics 2003;4:43–53. [PubMed: 14604442]
19. Milona MA, Gough JE, Edgar AJ. Genomic structure and cloning of two transcript isoforms of human Sp8. BMC Genomics 2004;5:86–97. [PubMed: 15533246]
20. Lu X, Gilbert L, He X, Rubin J, Nanes MS. Transcriptional regulation of the osterix (Osx, Sp7) promoter by tumor necrosis factor identifies disparate effects of mitogen-activated protein kinase and NF kappa B pathways. J Biol Chem 2006;281:6297–6306. [PubMed: 16410254]
21. Marin M, Karis A, Visser P, Grosveld F, Philipsen S. Transcription factor Sp1 is essential for early embryonic development but dispensible for cell growth and differentiation. Cell 1997;89:619–628. [PubMed: 9160753]
22. Bouwman P, Göllner H, Elsäaaer HP, Eckhoff G, Karis A, Grosveld F, Philipsen S, Suske G. Transcription factor Sp3 is essential for post-natal survival and late tooth development. EMBO J 2000;19:655–661. [PubMed: 10675334]
23. Van Loo PF, Bouwman P, Ling KW, Middendorp S, Suske G, Grosveld F, Dzierzak E, Philipsen S, Hendriks RW. Impaired hematopoiesis in mice lacking the transcription factor Sp3. Blood 2003;102:858–866. [PubMed: 12676787]
24. Göllner H, Dani C, Phillips B, Philipsen S, Suske G. Impaired ossification in mice lacking the transcription factor Sp3. Mech Devel 2001;106:77–83. [PubMed: 11472836]
25. Göllner H, Bouwman P, Mangold M, Karis A, Braun H, Rohner I, Del Rey A, Besedovsky HO, Meinhardt A, van den Broek M, Cutforht T, Grosveld F, Philipsen S, Suske G. Complex phenotype of mice homozygous for a null mutation in the Sp4 transcription factor gene. Genes Cells 2001;6:689–697. [PubMed: 11532028]
26. Nguyen-Tran VT, Kubalak SW, Minamisawa S, Fiset C, Wollert KC, Brown AB, Ruiz-Lozano P, Barrere-Lemaire S, Kondo R, Norman LW, Gourdie RG, Rahme MM, Feld GK, Clark RB, Giles WR, Chien KR. A novel genetic pathway for sudden cardiac death via defects in the transition between ventricular and conduction system cell lineages. Cell 2000;102:671–682. [PubMed: 11007485]
27. Moorefield KS, Fry SJ, Horowitz JM. Sp2 DNA binding activity and *trans*-activation are negatively regulated in mammalian cells. J Biol Chem 2004;279:13911–13924. [PubMed: 14726517]
28. Phan D, Cheng CJ, Galfione M, Vakar-Lopez F, Tunstead J, Thompson NE, Burgess RR, Najjar SM, Yu-Lee LY, Lin SH. Identification of Sp2 as a transcriptional repressor of carcinoembryonic antigen-related cell adhesion molecule 1 in tumorigenesis. Cancer Res 2004;64:3072–3078. [PubMed: 15126343]
29. Xie J, Nichols TD, Yoder JA, Horowitz JM. Sp2 is a maternally inherited transcription factor required for embryonic development. J Biol Chem. 2010 in press.
30. Simmons SO, Horowitz JM. Nkx3.1 binds and negatively regulates the transcriptional activity of Sp-family members in prostate-derived cells. Biochem J 2005;393:397–409. [PubMed: 16201967]
31. Darby IA, Bisucci T, Desmouliere A, Hewitson TD. In situ hybridization using cRNA probes: isotopic and nonisotopic detection methods. Methods Mol Biol 2006;326:17–31. [PubMed: 16780192]
32. Zurney J, Howard KE, Sherry B. Basal expression levels of IFNAR and Jak-STAT components are determinants of cell-type-specific differences in cardiac antiviral responses. J Virol 2007;81:13668–13680. [PubMed: 17942530]
33. Morales D, Hatten ME. Molecular markers of neuronal progenitors in the embryonic cerebellar anlage. J Neurosci 2006;26:12226–12236. [PubMed: 17122047]

34. Zervas M, Blaess S, Joyner AL. Classical embryological studies and modern genetic analysis of midbrain and cerebellum development. Curr Top Dev Biol 2005;69:101–138. [PubMed: 16243598]
35. Liu A, Joyner AL. Early anterior/posterior patterning of the midbrain and cerebellum. Annu Rev Neurosci 2001;24:869–896. [PubMed: 11520921]
36. Carninci P, Sandelin A, Lenhard B, Katayama S, Shimokawa K, Ponjavic J, Semple CA, Taylor MS, Engstrom PG, Frith MC, Forrest AR, Alkema WB, Tan SL, Plessy C, Kodzius R, Ravasi T, Kasukawa T, Fukuda S, Kanamori-Katayama M, Kitazume Y, Kawaji H, Kai C, Nakamura M, Konno H, Nakano K, Mottagui-Tabar S, Arner P, Chesi A, Gustincich S, Persichetti F, Suzuki H, Grimmond SM, Wells CA, Orlando V, Wahlestedt C, Liu ET, Harbers M, Kawai J, Bajic VB, Hume DA, Hayashizaki Y. Genome-wide analysis of mammalian promoter architecture and evolution. Nat Genet 2006;38:626–635. [PubMed: 16645617]
37. Young TL, Matsuda T, Cepko CL. The noncoding RNA taurine upregulated gene 1 is required for differentiation of the murine retina. Curr Biol 2005;15:501–512. [PubMed: 15797018]
38. Willingham AT, Orth AP, Batalov S, Peters EC, Wen BG, za-Blanc P, Hogenesch JB, Schultz PG. A strategy for probing the function of noncoding RNAs finds a repressor of NFAT. Science 2005;309:1570–1573. [PubMed: 16141075]
39. Ginger MR, Shore AN, Contreras A, Rijnkels M, Miller J, Gonzalez-Rimbau MF, Rosen JM. A noncoding RNA is a potential marker of cell fate during mammary gland development. Proc Natl Acad Sci U S A 2006;103:5781–5786. [PubMed: 16574773]
40. Caretti G, Schiltz RL, Dilworth FJ, Di PM, Zhao P, Ogryzko V, Fuller-Pace FV, Hoffman EP, Tapscott SJ, Sartorelli V. The RNA helicases p68/p72 and the noncoding RNA SRA are coregulators of MyoD and skeletal muscle differentiation. Dev Cell 2006;11:547–560. [PubMed: 17011493]
41. Feng J, Bi C, Clark BS, Mady R, Shah P, Kohtz JD. The Evf-2 noncoding RNA is transcribed from the Dlx-5/6 ultraconserved region and functions as a Dlx-2 transcriptional coactivator. Genes Dev 2006;20:1470–1484. [PubMed: 16705037]

Abbreviations

Sp	Specificity protein
Cys	Cysteine
His	Histidine
GC	Guanine and Cytosine
kb	kilobases
Mb	Megabases
HOX	Homeobox gene
DMEM	Dulbecco's modified Eagle's medium
PBS	phosphate-buffered saline
EDTA	ethylenediaminetetraacetic acid
PCR	Polyermase Chain Reaction
dATP	deoxyadenosine triphosphate
dI/dC	deoxyinosine/deoxycytosine
Leu	Leucine
Trp	Tryptophan
SD	Synthetic Dropout
mM	millimolar
SSC-DEPC	Standard Saline Citrate prepared with diethylpyrocarbonate-treated water

Fab-AP	Fragment antigen binding-Alkaline phosphate
NBT/BCIP	Nitro-Blue Tetrazolium Chloride/5-Bromo-4-Chloro-3′-Indolyphosphate p-Toluidine Salt
RT-PCR	Reverse Transcription-Polymerase Chain Reaction
5′RACE	Rapid Amplification of 5′ Complementary Ends
EGL	External Granule Layer
IGL	Internal Granule Layer
PCL	Purkinje Cell Layer
OLB	Olfactory Bulb
RMS	Rostral Migratory Stream
DG	Dentate Gyrus
SVZ	Subventricular Zone
ML	Molecular Layer
GCL	Granule Cell Layer
MCL	Mitral Cell Layer
GL	Glomerular Layer

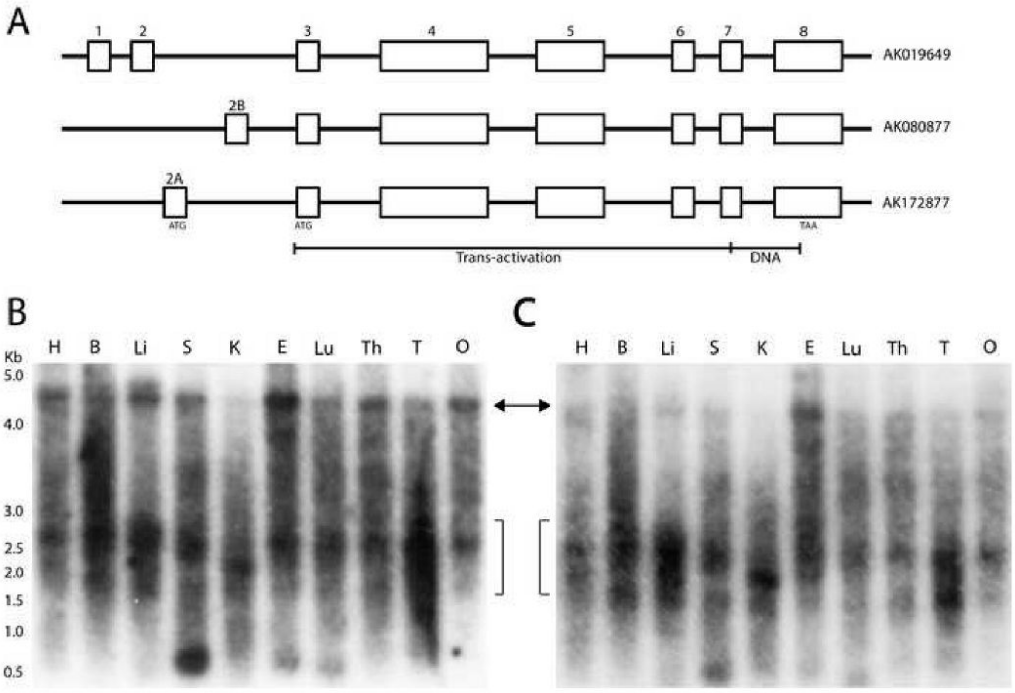

Fig. 1.

Transcription of the mouse Sp2 locus results in alternatively-spliced as well as sub-genomic-length messages. (A) Three classes of alternatively-spliced Sp2 messages. Alternative splicing upstream of the first presumed coding exon (exon 3) links exons encoding one of three distinct 5′ untranslated regions to Sp2 coding sequences. Exons 1 and 2 have been identified in a testis-derived cDNA (AK019649), exon 2B-containing cDNAs have been identified in a neonatal adipose tissue-derived cDNA (AK080877) as well as a cDNA derived from heart-derived fibroblasts in the present study. Exon 2A has been identified in a mixed-tissue-derived cDNA (AK172877) as well as a wide-variety of additional cDNAs including a heart-derived cDNA (GU126673) obtained in this study. Exons are indicated by numbered boxes, putative translational start sites in exons 2A and 3 are indicated as is the translational stop site in exon 8, and sequences encoding the Sp2 *trans*-activation and DNA-binding domains are also indicated. (B) Northern blot of poly-A(+) RNAs from mouse tissues analyzed with a radiolabeled probe derived from exon 3 and the 5′ end of exon 4 (probe 1). (C) Northern blot of poly-A(+) RNAs from mouse tissues analyzed with a radiolabeled probe derived from the 3′ end of exon 4, exon 5, and the 5′ end of exon 6 (probe 2). RNAs examined in (B) and (C) were obtained from the following tissues: H, heart; B, Brain; Li, Liver; S, Spleen; K, Kidney; E, Whole Embryo; Lu, Lung; Th, Thymus; T, Testis; O, Ovary. A double-headed arrow indicates full-length Sp2 messages and brackets indicate sub-genomic Sp2-derived messages. Molecular weight markers are indicated on the left.

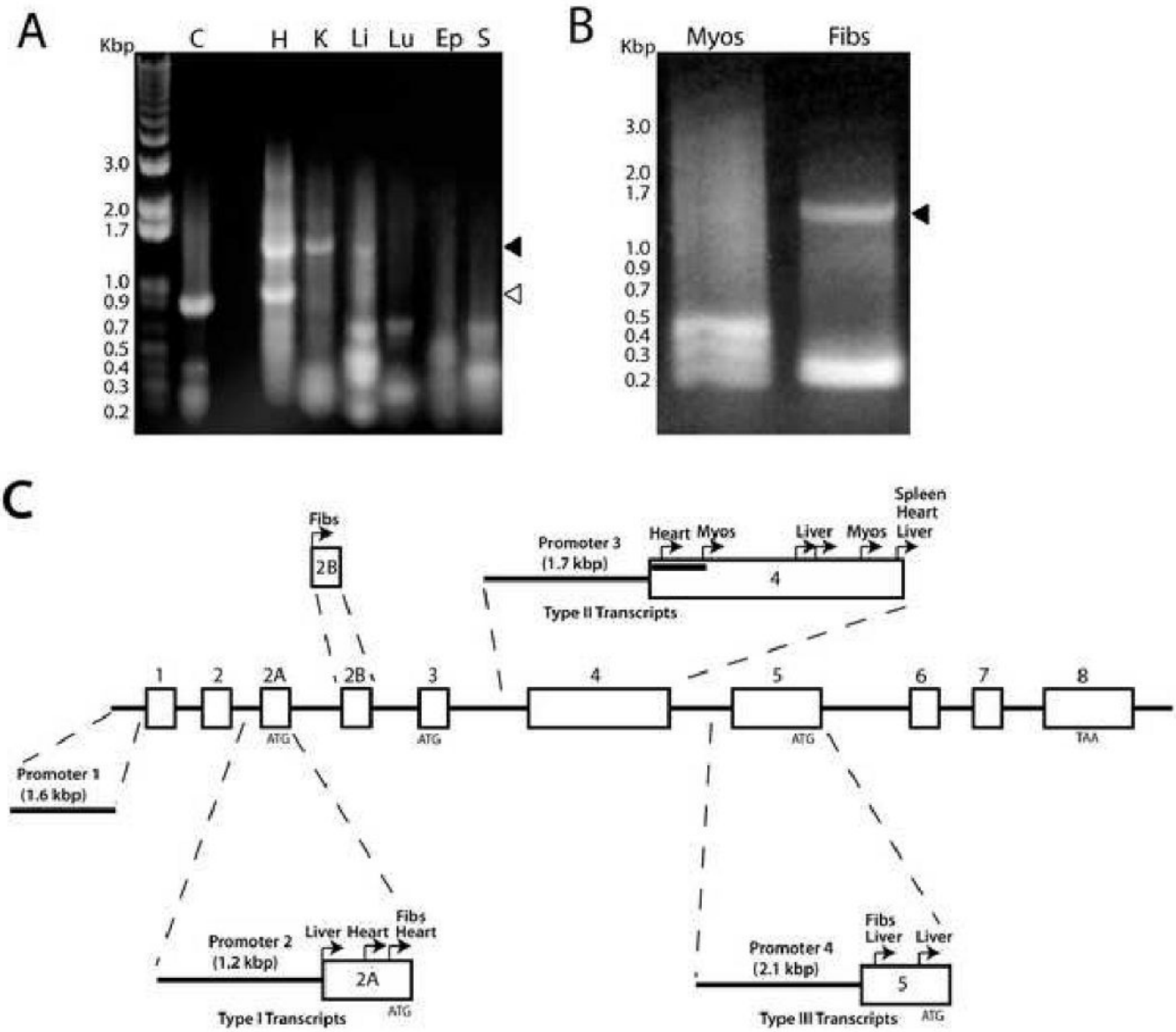

Fig. 2.

Tissue- and cell type-specific Sp2 transcripts and localization of sites of transcriptional initiation via 5′RACE. (A) 5′RACE analysis of RNAs harvested from mouse tissues. Nested anti-sense primers that bound sequences within Sp2 exons 5 and 7 were employed to synthesize 5′RACE products using total RNAs from the following tissues: H, Heart; K, Kidney; Li, Liver, Lu, Lung, Ep, Epidermis, S, Spleen. 5′RACE products (ca. 1380 bp) that result from transcriptional initiation within exon 2A are indicated by a filled arrowhead. A heart-specific 895 bp 5′RACE product that results from transcriptional initiation at the 5′ end of exon 4 is indicated by an open arrowhead. A control 5′RACE reaction, in lane labeled C, utilized HeLa cell RNA and gene-specific primers to amplify an 827 bp fragment derived from the 5′ end of human β-actin. Molecular weight markers are indicated on the left. (B) 5′RACE analysis of RNAs harvested from mouse heart-derived cell cultures. Total RNAs prepared from heart myocyte (Myos) and heart fibroblast (Fibs) cultures were analyzed by 5′RACE as in (A). A closed arrow indicates transcripts (ca. 1350 bp) that initiate within exons 2A or 2B. Molecular weight markers are indicated on the left. (C) Schematic diagram indicating transcriptional

initiation sites within the mouse Sp2 locus as determined by 5′RACE analysis as well as genomic fragments assayed for transcriptional activity in transient transfection experiments. Type I, II, and III transcripts result from transcriptional initiation within exons 2A, 4, or 5 respectively. Mapped sites of transcriptional initiation are indicated above each corresponding exon as well as the tissue or cell in which they were detected. The initiation site of heart fibroblast-derived 5′RACE product in exon 2B is also indicated. Genomic fragments (Promoters 1, 2, 3, and 4) assayed for promoter activity in transient transfection experiments are indicated as is a fragment of the 5′ end of exon 4 (Probe 3; black bar) that was employed as a molecular probe in *in situ* hybridization experiments. Numbered exons and translational start and stop sites are indicated as in Fig. 1.

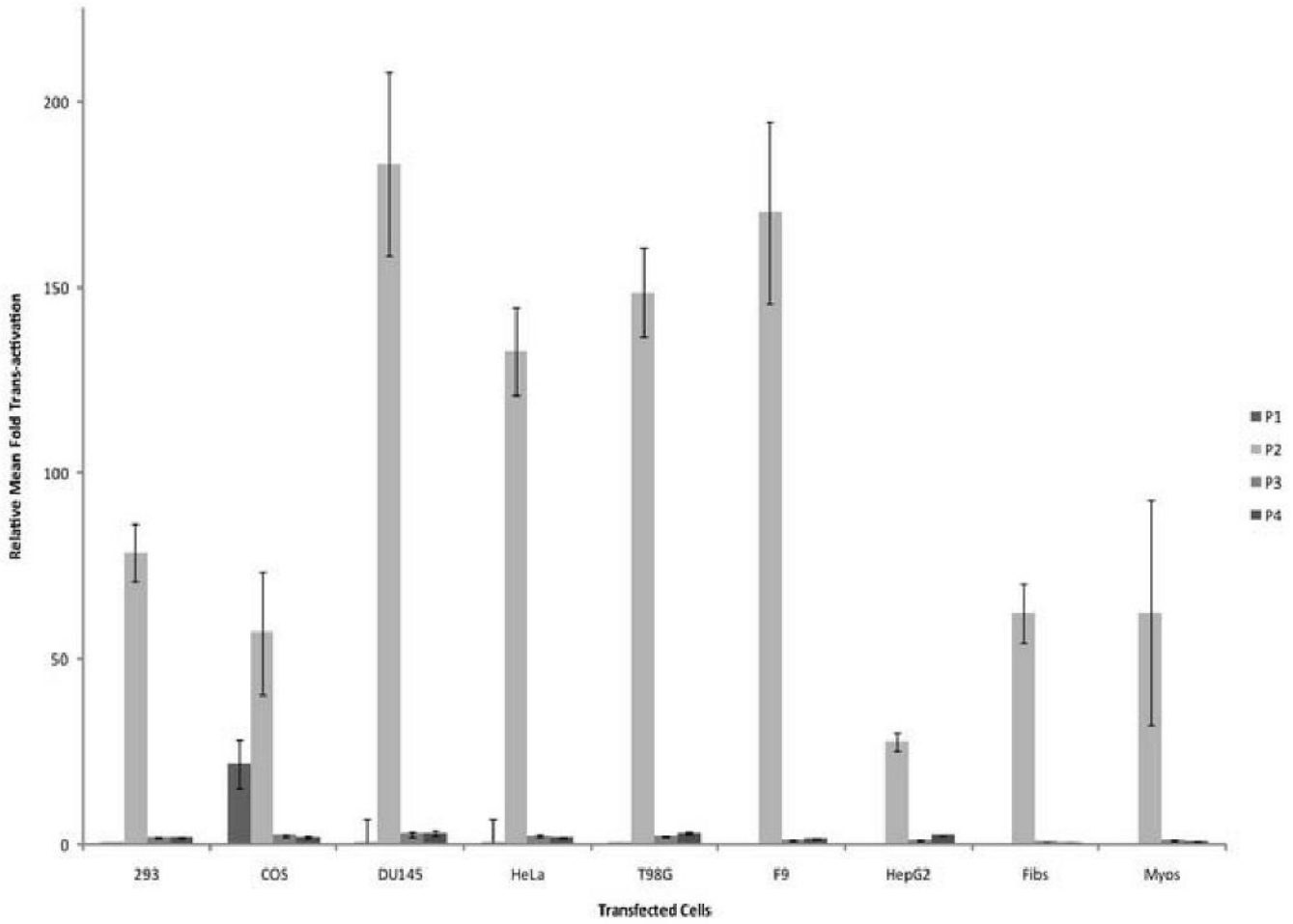

Fig. 3.

Transcriptional activities of genomic fragments derived from the mouse Sp2 locus in nine mammalian cell lines and cell strains. Genomic fragments corresponding to Promoters 1-4 were sub-cloned in pGL3-basic generating pGL3-P1, pGL3-P2, pGL3-P3, and pGL3-P4. Firefly luciferase activities elicited by each reporter construct in each cell type examined were normalized to a co-transfected *Renilla* luciferase reporter construct (Δ53RL), and levels of mean fold *trans*-activation were calculated relative to luciferase activity produced by pGL3-basic (set equal to one). Relative levels of mean fold *trans*-activation are indicated with error bars corresponding to the standard error of the mean. Data presented is derived from at least three independent transfections of each cell type examined.

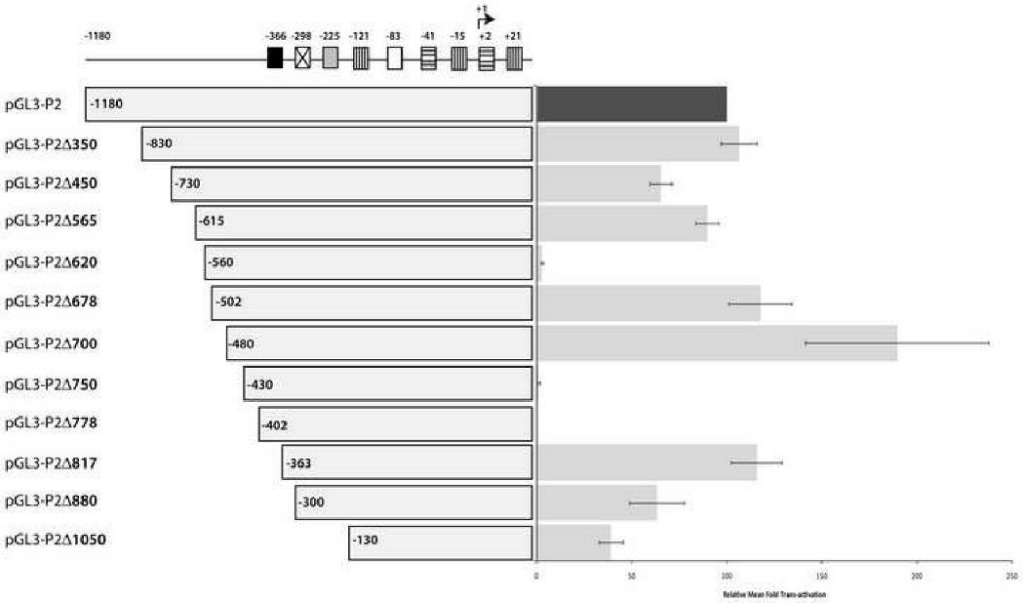

Fig 4.

Transcriptional activities of nested deletions prepared from Sp2 Promoter 2. The transcriptional activities of pGL3-P2 and a series of nested deletions were determined in the nine mammalian cell lines and cell strains as indicated in Fig. 3. Relative levels of mean fold *trans*-activation in all cell types examined are indicated with error bars corresponding to the standard error of the mean. Data presented is derived from at least three independent transfections of each cell type examined. A schematic diagram of the 1.2 kbp genomic fragment carried by pGL3-P2 is indicated at the top as well as nine evolutionarily-conserved putative transcription factor binding sites (boxes) described in the text: black box, CDP/Cut; crossed box, LTR-derived CCAAT; gray box, Sp2; vertical hatched boxes, CCAAT/NF-Y; open box, CCAAT; horizontal hatched boxes, GC-rich. The 5′ endpoints of each deletion construct are indicated relative to the most 5′ site of type II transcriptional initiation (+1; indicated by an arrow).

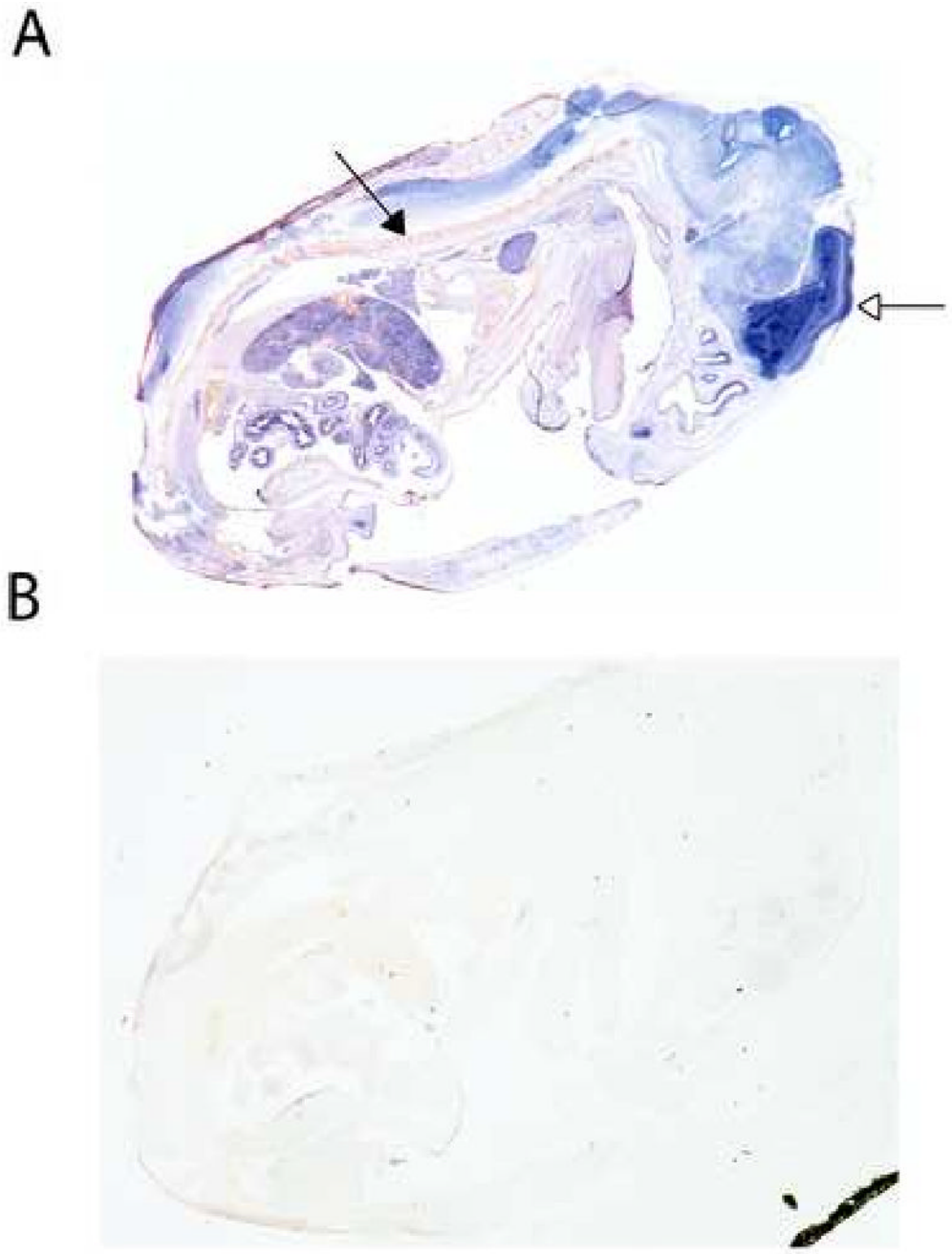

Fig. 5.

Detection of Sp2 transcripts in mouse embryo sections by *in situ* hybridization. Anti-sense (A) and sense (B) digoxigenin-labeled RNA probes were prepared from a 489 bp portion of the 5′ end of mouse Sp2 exon 4 (Probe 3). Each probe was hybridized to sagittal sections prepared from mouse embryos at day 13 (E13) of gestation. Arrows indicated the developing forebrain (open arrowhead) and spine (closed arrowhead).

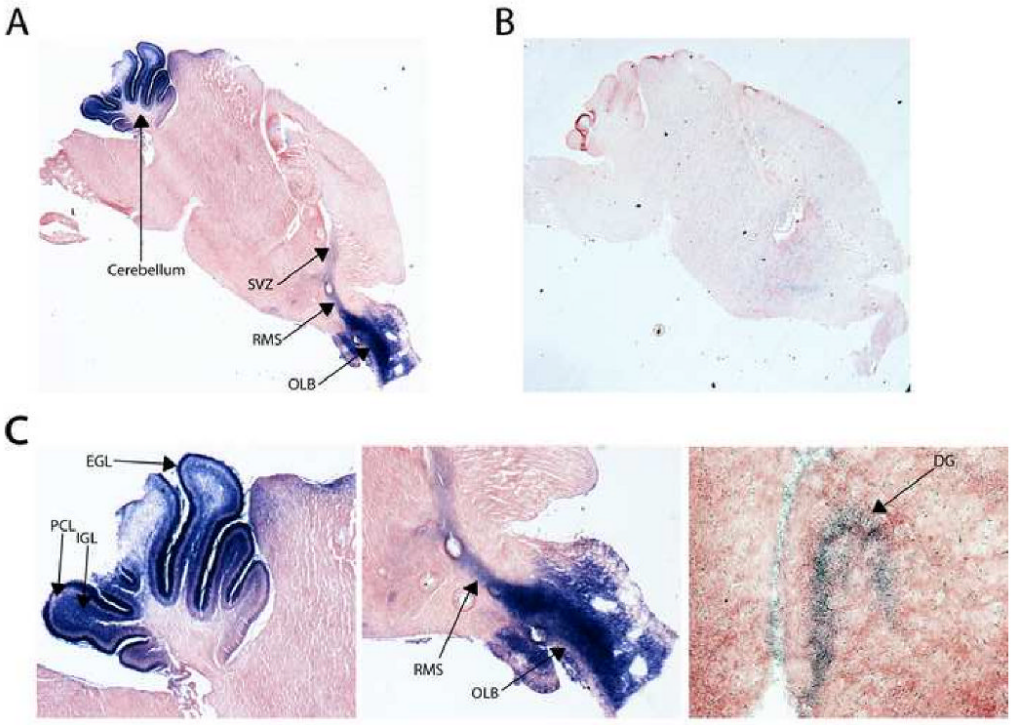

Fig. 6.

Detection of Sp2 transcripts in post-natal mouse brain sections by *in situ* hybridization. (A) Sagittal sections of post-natal day 5 mouse brain were hybridized with anti-sense (A and C) or sense (B) digoxigenin-labeled probes as in Fig. 5. Hybridized portions of mouse brain in (A) include the cerebellum, subventricular zone (SVZ), rostral-migratory stream (RMS), and olfactory bulb (OLB). Higher power images of the section shown in (A), appear in the panels in (C). (C) Indicated are the external (EGL) and internal (IGL) granular layers and the Purkinje cell layer (PCL) of the cerebellum (left panel), the rostral-migratory stream (RMS) and olfactory bulb (OLB; center panel), and the dentate gyrus (DG) of the hippocampus (right panel).

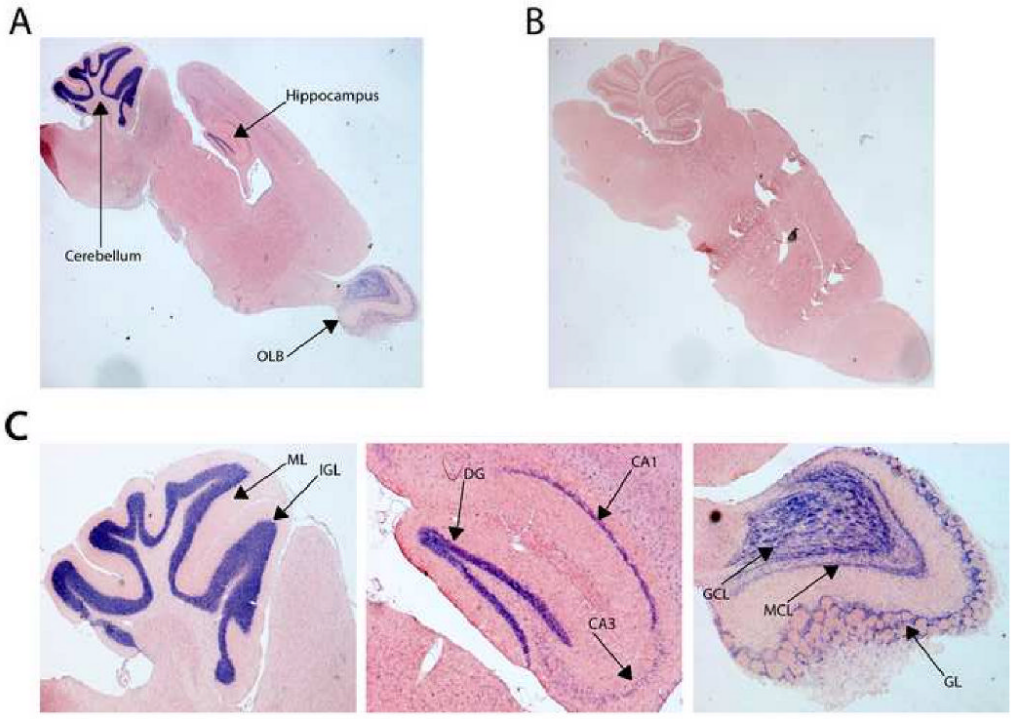

Fig. 7.

Detection of Sp2 transcripts in adult mouse brain sections by *in situ* hybridization. (A) Sagittal sections of the brain of a mouse at three months of age were hybridized with anti-sense (A and C) or sense (B) digoxigenin-labeled probes as in Fig. 5. Higher power images of the section shown in (A), appear in the panels in (C). (C) Indicated are the molecular (ML) and internal granule layers (IGL) of the cerebellum (left panel), the dentate gyrus (DG) and CA1 and CA3 layers of the hippocampus (center panel), and the granule cell (GCL), mitral cell (MCL), and glomerular layers (GL) of the olfactory bulb (right panel).