

# Identification of evolutionarily conserved, functional noncoding elements in the promoter region of the sodium channel gene *SCN8A*

Valerie L. Drews · Kehui Shi · Georgius de Haan ·  
Miriam H. Meisler

Received: 23 February 2007 / Accepted: 29 July 2007 / Published online: 9 October 2007  
© The Author(s) 2007

**Abstract** *SCN8A* is a major neuronal sodium channel gene expressed throughout the central and peripheral nervous systems. Mutations of *SCN8A* result in movement disorders and impaired cognition. To investigate the basis for the tissue-specific expression of *SCN8A*, we located conserved, potentially regulatory sequences in the human, mouse, chicken, and fish genes by 5' RACE of brain RNA and genomic sequence comparison. A highly conserved 5' noncoding exon, exon 1c, is present in vertebrates from fish to mammals and appears to define the ancestral promoter region. The distance from exon 1c to the first coding exon increased tenfold during vertebrate evolution, largely by insertion of repetitive elements. The mammalian gene acquired three novel, mutually exclusive noncoding exons that are not represented in the lower vertebrates. Within the

shared exon 1c, we identified four short sequence elements of 10–20 bp with an unusually high level of evolutionary conservation. The conserved elements are most similar to consensus sites for the transcription factors Pou6f1/Brn5, YY1, and REST/NRSF. Introduction of mutations into the predicted Pou6f1 and REST sites reduced promoter activity in transfected neuronal cells. A 470-bp promoter fragment containing all of the conserved elements directed brain-specific expression of the LacZ reporter in transgenic mice. Transgene expression was highest in hippocampal neurons and cerebellar Purkinje cells, consistent with the expression of the endogenous gene. The compact cluster of conserved regulatory elements in *SCN8A* provides a useful target for molecular analysis of neuronal gene expression.

---

Valerie L. Drews, Kehui Shi contributed equally to this work.

---

**Electronic supplementary material** The online version of this article (doi:10.1007/s00335-007-9059-8) contains supplementary material, which is available to authorized users.

---

V. L. Drews · K. Shi · G. de Haan · M. H. Meisler  
Department of Human Genetics, University of Michigan,  
Ann Arbor, Michigan 48109-0618, USA

*Present Address:*

V. L. Drews  
Department of Medicine, UCSF, San Francisco, CA, USA

*Present Address:*

G. de Haan  
Broad Institute, Boston, MA, USA

M. H. Meisler (✉)

Department of Human Genetics, University of Michigan,  
4909 Buhl, Box 0618, Ann Arbor, MI 48109-0618, USA  
e-mail: meislerm@umich.edu

## Introduction

The sodium channel gene *SCN8A* encodes Na<sub>v</sub>1.6, one of the major voltage-gated sodium channels responsible for generating action potentials during neuronal firing. *SCN8A* is expressed throughout the central and peripheral nervous systems, with particularly high expression in cerebellum, and is the major sodium channel at adult nodes of Ranvier (Caldwell et al. 2000; Schaller et al. 1995). Mutations of *SCN8A* are associated with neurologic disorders in human and mouse. Missense and nonsense mutations of mouse *Scn8a* result in movement disorders, including tremor, ataxia, dystonia, and paralysis (Burgess et al. 1995; Meisler et al. 2004). Conditional knockout of mouse *Scn8a* in Purkinje cells is sufficient to generate tremor and ataxia (Levin et al. 2006). Heterozygotes for a null allele of human *SCN8A* exhibit ataxia and/or impaired cognition (Trudeau et al. 2006). *SCN8A* orthologs with expression in brain

have been identified in several species of fish (Lopreato et al. 2001; Novak et al. 2006).

The transcriptional regulation of *SCN8A* is not well characterized. We recently described the promoter of the mouse and human *SCN8A* genes, which contain a cluster of four mutually exclusive 5' noncoding exons, exon 1a to 1d, each of which is spliced directly to the first coding exon (Drews et al. 2005). A 4.8-kb genomic fragment containing all of the noncoding exons demonstrated tissue-specific expression in transgenic mice (Drews et al. 2005). An 0.85-kb subfragment containing exon 1b and exon 1c was expressed at a high level in a neuronal cell line.

We now report the use of evolutionary sequence comparison to identify highly conserved noncoding sequences in the promoter region of *SCN8A*. Large-scale comparisons between mammalian, chicken, and fish genomes have been used to identify highly conserved regulatory regions that are several hundred nucleotides in length (Dodou et al. 2003; Ghanem et al. 2003; Nobrega et al. 2003; Spitz et al. 2003; Zerucha et al. 2000). Genomic sequence comparisons alone are not sufficient for identifying short conserved elements, 10–20 bp in length, the typical length of binding sites for regulatory proteins. To bypass this limitation, we performed a two-step approach: (1) 5' RACE to identify the start sites of transcription initiation in chicken, fish, and mammals, and (2) sequence comparison of the surrounding genomic DNA. This approach enabled us to identify a cluster of four short noncoding elements that are largely invariant in vertebrate genomes. In this article we describe the sequence and preliminary functional characterization of these evolutionarily conserved noncoding elements.

## Materials and methods

### 5' RACE

5' RACE was performed using the GeneRacer Kit (Invitrogen, Carlsbad, CA) with 250 ng of chicken brain poly(A)<sup>+</sup> RNA (Ambion, Austin, TX) as template and reverse primers complementary to the first coding exon of *SCN8A* (exon 1). Three primers were used in succession: primer 1 for reverse transcription (5'GGTTT GCTGT CTTCA TCGTC GTC), primer 2 for PCR (5'TCTGC AATGC GTTTC TCAAT GTTAG), and primer 3 for nested PCR (5'AGCCG TGCTG CCATC TTTTC ATC). PCR amplification was initiated by 2 min of denaturing at 94°C followed by 33 cycles of 30 sec at 94°C, 30 sec at 65°C, and 1 min at 72°C, with a final extension step of 6 min at 72°C. PCR products were cloned into the pGEMT-Easy vector (Promega, Madison, WI) using the Quick Ligase Kit (New England BioLabs, Ipswich, MA). Inserts were amplified by PCR and visualized by ethidium bromide staining of agarose gels. Inserts of unique size were purified

using the QIAquick Gel Extraction Kit (Qiagen, Valencia, CA) and sequenced at the University of Michigan Sequencing Core (<http://www.seqcore.brcf.med.umich.edu/>).

### Multispecies DNA sequence analysis

*SCN8A* genomic sequences from the following sources were used: *Homo sapiens* chromosome 12 BAC clone RP11-285E4 (GenBank AC025097), *Mus musculus* chromosome 15 BAC clone RP23-319B16 from strain C57BL/6J (GenBank AC104833), *Fugu rubripes* whole-genome shotgun sequence SCAFFOLD 1918 (GenBank CAAB01001918.1), *Monodelphis domestica* (opossum) whole-genome shotgun sequence (GenBank AAFR03066481), and *Gallus gallus* whole-genome shotgun sequence assembly (GenBank NW060828). The genomic sequence for the two duplicated *Scn8a* genes in *Danio rerio* was obtained from clone DKEY-9P24 (GenBank CR376824.2) (*Scn8aa*) and RP71-15H20 (GenBank BX470135.6) (*Scn8ab*). Mouse exon 1c sequence was reported by Drews et al. (2005) (GenBank AY510081). Human exon 1c sequence was obtained from mRNA sequence (GenBank NM014191), and chicken exon 1c sequence was determined by 5' RACE (GenBank EF210713). Zebrafish exon 1c sequence was obtained from a *Scn8ab* mRNA (GenBank NM131628), and subsequent alignment of that sequence to genomic sequence upstream of *Scn8aa* coding sequence (GenBank NM001045183). Sequences were aligned using Sequencher software (GeneCodes, Ann Arbor, MI). MatInspector was used to identify potential transcription factor binding sites (<http://www.genomatix.de/products/MatInspector/index.html>). The repeat content of human genomic DNA was analyzed using RepeatMasker ([www.repeatmasker.org](http://www.repeatmasker.org)) and PipMaker (<http://www.bio.cse.psu.edu/pipmaker/>). The pictogram of conserved sequence elements was assembled using Pictogram software (<http://www.genes.mit.edu/pictogram.html>).

### Luciferase constructs

The 470-bp *Scn8a* promoter-luciferase construct, p470Luc (Fig. 3, top) was constructed from the previously described 0.85-kb promoter construct (construct 6, Drews et al. 2005) by digestion with *AscI* and *MluI* followed by religation. Mutations of the conserved elements were introduced to this construct using two synthetic 90mers containing the restriction sites for *Eco47III* and *BglII* at their termini (Sigma, St. Louis, MO). After digestion of the wild-type construct with these two enzymes, the wild-type fragment was replaced with the synthetic fragments. In the mutated construct pGL3ΔB, the conserved sequence CAAGATGGCG in element B was changed to

**Table 1** Transcript sequence data for the 5' noncoding exon 1c of *SCN8A* from five vertebrate species

Species	Type	GenBank	bp	% ID to human	Av. ID of 5' UTR	Observed splice donor site
Human	mRNA	NM_014191	124	100%	100%	CCGCTGCCgtaagtcg
Mouse	5'-RACE	AY510081	42	93%	70%	CCGCCGCCgtaagttg
Chicken	5'-RACE	EF210713	176	72%	2%	CGGCCGCAgtaagtga
Zebrafish	mRNA	NM_131628	229	47%	–	CAATCATAgtaagtat

The percent sequence identity for *SCN8A* is much higher than the average values for 5' UTRs in these genomes (ICGS 2004; Makalowski and Boguski 1998). The zebrafish data are for *Scn8ab*

CGGAACCGAG. In the mutated construct pGL3 $\Delta$ DR, the first copy of the conserved repeat GCAGT in the direct repeat is changed to TACGG and the second copy is changed to TGACG. All constructs were verified by sequencing.

#### Cell culture and transfection

The mouse neuronal hybrid cell line MN-1 was cultured and transfected as previously described (Drews et al. 2005), using 50 ng of test plasmid, 10 ng of control plasmid (pRL-SV40, Promega), and the Fugene 6 transfection reagent (Roche Molecular Biochemicals, Indianapolis, IN). Cell lysates were collected at 40–48 h post-transfection. Firefly luciferase and Renilla luciferase were assayed using the Dual-Luciferase Reporter Assay System (Promega). Each transfection was performed in triplicate; each construct was analyzed in two to four independent transfection experiments. The mean and standard error for luciferase activity were calculated using the Statistical Package for the Social Sciences (SPSS Inc., Chicago, IL).

#### Transgenic mice

The wild-type plasmid p470Luc was digested with *Xba*I and *Nco*I, and the luciferase cDNA insert was replaced with the nLacZ reporter fragment from Hsp68-LacZ (Nobrega et al. 2003) to generate the transgene p470LacZ. The linearized insert was gel purified and microinjected into (C57BL/6J  $\times$  SJJL) $F_2$  fertilized eggs in the Transgenic Animal Model core at the University of Michigan (<http://www.med.umich.edu/tamc/>). Mice carrying the transgene were identified by PCR with the forward primer 5'CTAAC GAAGC TGCTG CAGAA TGAG from the *Scn8a* promoter and the reverse primer 5'GTTTG AGGGG ACGAC GACAG TATC from LacZ. Transgenic founders were crossed with strain C57BL/6J to generate  $F_1$  mice for analysis of LacZ expression. Brain, kidney, liver, and spleen were frozen, sectioned, and stained for  $\beta$ -galactosidase activity with the Xgal substrate (Histoserv, Germantown, MD).

## Results

Chicken *Scn8a* transcripts contain a single 5' noncoding exon related to mammalian exon 1c

To identify the transcription start site for chicken *Scn8a*, we performed 5' RACE on poly(A)<sup>+</sup> brain RNA, using reverse primers located in the first coding exon. Forty-three clones with inserts of 13 different size classes were analyzed. Sequencing revealed that all of the clones contained the same 5' noncoding exon, which was spliced directly to the first coding exon (Table 1). The chicken noncoding exon ends with a consensus splice donor site. The longest clone contained 175 bp of upstream noncoding sequence. Sequence comparison of the chicken noncoding exon with the mammalian exons 1a-1d demonstrated 72% sequence identity (74/103 bp) with mammalian exon 1c. This degree of conservation is unusual for 5' UTR sequences, which exhibit on average only 2% conservation in chicken and mammalian orthologs (ICGS 2004, Fig. 2). The identification of a single chicken 5'-RACE product contrasts with our previous identification of four different noncoding exons by 5' RACE of human and mouse transcripts (Drews et al. 2005), and suggests that mammalian exon 1c represents the ancestral exon. No matches to the other mammalian noncoding exons were detected in adjacent chicken genomic DNA sequence.

Fish *Scn8a* contains a 5' noncoding exon related to mammalian exon 1c

Sequence comparison between the chicken noncoding exon and fugu and zebrafish genomic sequences upstream of the first coding exon of *Scn8a* identified a fish homolog of exon 1c, which was also present in one zebrafish cDNA clone (Table 1). The fish exon terminates in a consensus splice donor site that has diverged by a few base pairs from the mammalian exon (Table 1; Fig. 1, top panel). The identification of an ortholog of exon 1c in the fish further supports the view that exon 1c encodes the ancestral 5' UTR. The 47% sequence identity between exon 1c in fish and mammals is unusually high for noncoding sequence,

suggesting that there has been selection to retain functional elements within this noncoding exon. There are no matches to the other mammalian noncoding exons in the fish genomic sequence.

Four short, highly conserved noncoding elements in exon 1c

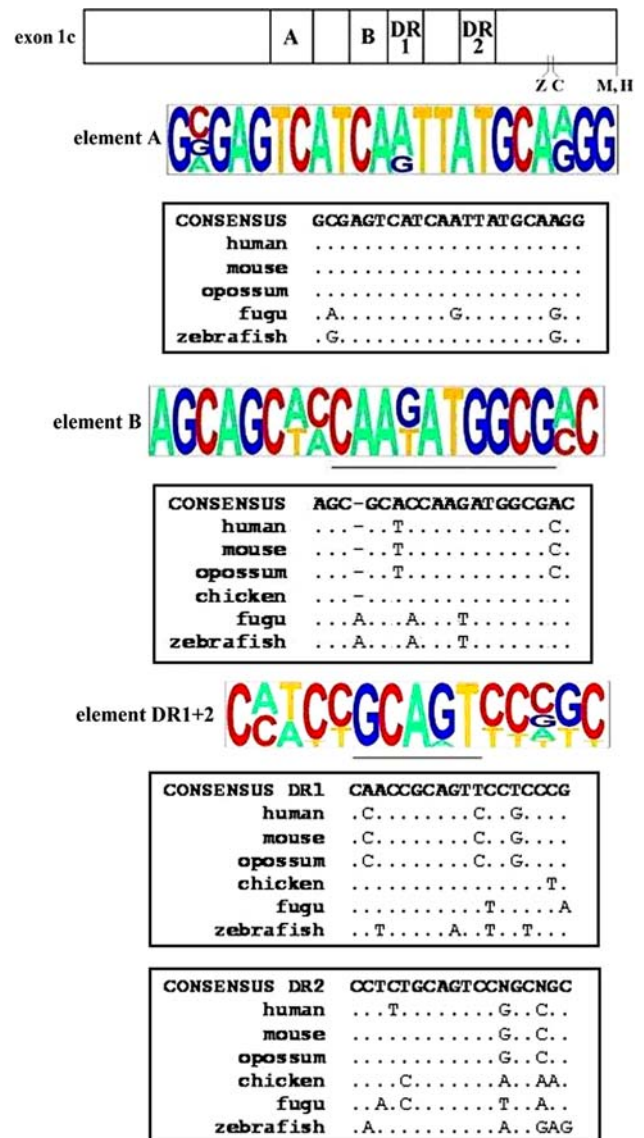
Sequence alignment of exon 1c from human, mouse, rat, opossum, chicken, zebrafish, and fugu revealed four regions of particularly high conservation. Element A is a 22-bp sequence located near the 5' end of exon 1c, with 20/22-bp sequence identity between mammals and fish (Fig. 1). (This element is not represented in the chicken gene.) Element B is a 20-bp sequence with 15/20-bp sequence identity between human, fish, and chicken (Fig. 1). Two direct repeats of 18 bp, designated DR1 and DR2, are separated by 14 bp in mammalian *SCN8A*. DR1 and DR2 are highly conserved in all five species examined (Fig. 1). Sequence conservation was not detected in genomic sequence immediately upstream and downstream of exon 1c (Supplementary Fig. 1).

#### Genomic organization of the 5' region of *SCN8A*

Mouse exon 1c and human exon 1c are located 65 kb and 70 kb upstream of the first coding exon, respectively (Drews et al. 2005). To determine the location of exon 1c in chicken and fish, we aligned the exon sequence with the genomic sequence. The chicken exon aligned to the 5' end of a gap in the genomic sequence (GenBank NW060828). We carried out PCR with primers upstream and downstream of the gap provided the missing sequence, which was only 100 bp long. This located the chicken exon 1c to a position 9 kb upstream of the first coding exon. In the two zebrafish *Scn8a* genes, the noncoding exon is located 9 kb and 14 kb upstream of the first coding exon, respectively. In the fugu genome, the noncoding exon is 3 kb upstream of exon 1 (Fig. 2).

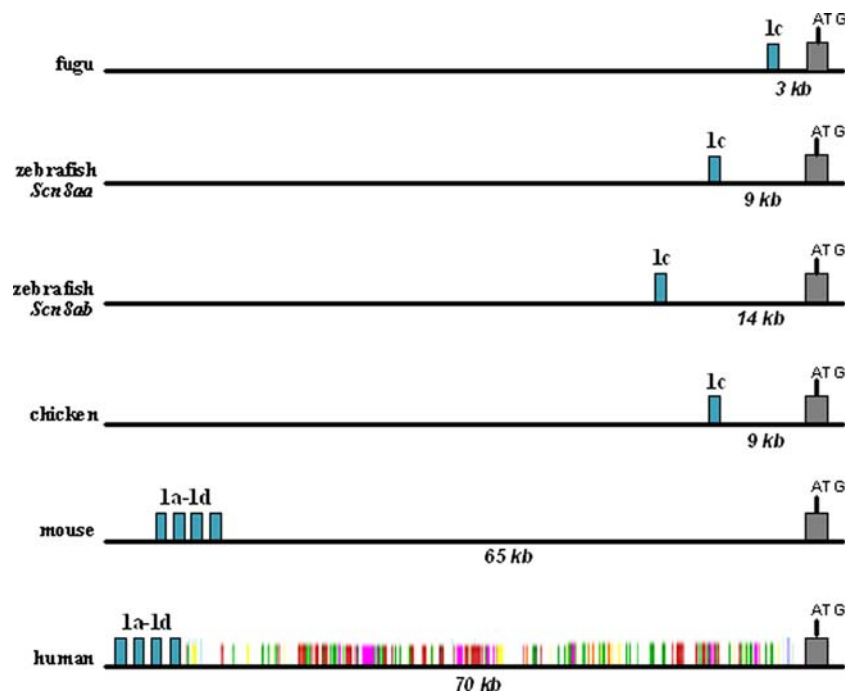
In the human gene, repetitive elements and simple sequences recognized by the RepeatMasker program account for 58% of the 70-kb interval between the 5' noncoding exons and the first coding exon (Fig. 2, bottom panel). The increase in size of the *SCN8A* gene thus appears to have occurred largely by invasion of repetitive elements, as is characteristic of the genome as a whole.

Exon 1c appears to be the only 5' noncoding exon in chicken and fish, based on our 5'-RACE data for chicken as well as fish ESTs. There is no evidence for conservation of mammalian exons 1a, 1b, and 1d, or substitution of other 5' exons, from genomic or transcript sequences. Exon 1a



**Fig. 1** Alignment of four conserved noncoding elements in the 5' UTR of vertebrate *SCN8A* genes. Relative positions within exon 1c are shown in the top panel. The variation in position of the splice donor sites (SD) in these species is indicated below the exon (z, zebrafish *Scn8a*; o, opossum; c, chicken; m, mouse; h, human). The YY1 site in the promoter is present in reverse orientation from that shown. Ten copies of the direct repeat were combined to generate the DR1&DR2 consensus, 5 from DR1 in each species and 5 from DR2. In the pictogram representation of consensus elements, the height of each letter represents the nucleotide frequency (<http://www.genes.mit.edu/pictogram.html>)

overlaps with an inserted LINE element in the human and mouse genes; the absence of a LINE element at this position in the opossum gene indicates a time of LINE insertion between 80 and 135 million years ago. Transcripts containing the four noncoding exons are expressed in multiple regions of the mouse brain and nervous system, indicating that they do not determine distinct expression domains (Drews et al. 2005).



**Fig. 2** Evolution of gene organization in the 5' region of *SCN8A*. The length of the interval between the 5' noncoding exon and the first coding exon increased from 3 kb to 70 kb during vertebrate evolution. In mammalian *SCN8A*, this interval is largely occupied by repetitive elements (bottom panel). After divergence from chicken, the mammalian lineage acquired three additional 5' noncoding exons that lack

sequence similarity to exon 1c. Exon 1a is derived from LINE element sequence (Drews et al. 2005). The four mammalian 5' noncoding exons are mutually exclusive; each is spliced directly to the first coding sequence (Drews et al. 2005). Repeated sequences recognized by RepeatMasker software are shown in color in the bottom panel. Exons not drawn to scale

#### Mutation of conserved elements reduces promoter activity in transfected cells

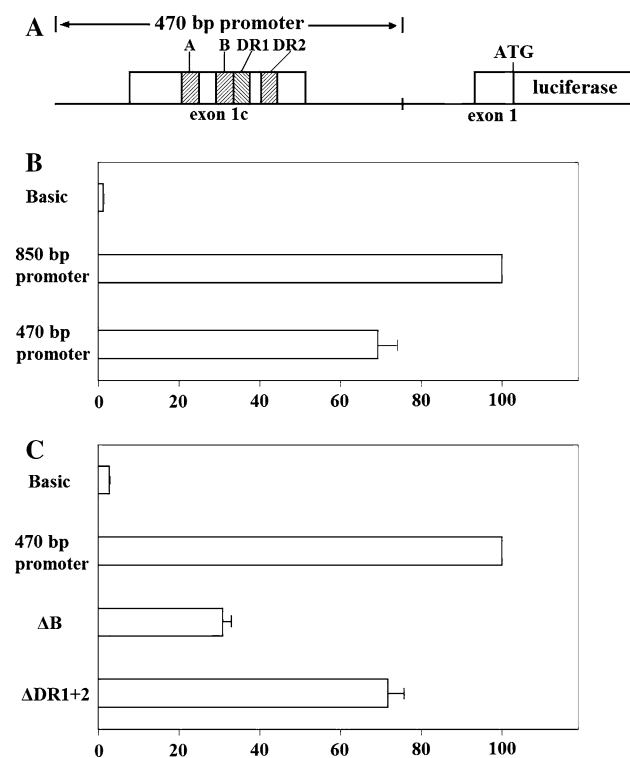
In previous studies of the mouse *Scn8a* promoter, we described an 850-bp fragment with strong promoter activity in transfected MN1 neuronal cells (Drews et al. 2005). To further localize the active promoter, we tested a 470-bp subfragment containing exon 1c as the only noncoding exon (Fig. 3A). This fragment retained 70% of the activity of the 850-bp fragment (Fig. 3B). To test the role of conserved elements B and DR1/DR2 in reporter gene expression, we generated two mutant constructs. In the  $\Delta B$  construct, the ten most conserved nucleotides of element B were changed from CAAGATGGCT to TGGAACCGAG, without changing GC content or introducing new transcription factor binding sites detectable with MatInspector software. This mutation resulted in a 70% reduction in luciferase expression (Fig. 3C). In the second mutant construct,  $\Delta DR1\&DR2$ , the invariant 5-bp core sequence of the direct repeat GCAGT was changed to TACGG and TGACG in DR1 and DR2, respectively.  $\Delta DR1\&DR2$  exhibited 30% reduction in expression compared with the wild-type fragment (Fig. 3C). Both element B and the direct repeat thus exhibit transcriptional enhancer activity in transfected cells.

#### Brain-specific expression in transgenic mice

The transgene p470LacZ containing the 470-bp promoter fragment directing expression of the nuclear LacZ reporter was analyzed in three independent lines of transgenic mice.  $\beta$ -Galactosidase activity was present throughout the brain in all three lines, with particularly high expression in hippocampal neurons and cerebellar Purkinje cells (Fig. 4). Expression was also high in the cortex, thalamus, and hypothalamus, with less intense staining visible throughout the brain. There was no activity in the olfactory bulb in any of the three lines, although the endogenous gene and the 4.8-kb transgene are strongly expressed there (Drews et al. 2005). An olfactory-specific enhancer thus appears to be missing from the 470-bp construct. No activity was detected in kidney, spleen, and liver of any of the lines (Fig. 4; data not shown). With the exception of the olfactory bulb, the pattern of expression of p470Luc reflects the tissue specificity of the endogenous *Scn8a* gene.

#### *In silico* analysis of transcription factor binding sites in the conserved noncoding elements of exon 1c

We used the MatInspector program for *in silico* analysis of the conserved elements in exon 1c to detect similarity



**Fig. 3** Transcription-enhancing activity of conserved element B and direct repeats 1 and 2. **A** The three-part expression construct p470Luc contains a 470-bp promoter fragment including the noncoding exon 1c, a 150-bp fragment with the splice acceptor and start codon from *SCN8A*, and the luciferase cDNA. **B** Firefly luciferase activity was normalized to cotransfected Renilla luciferase. Luciferase activity is presented as percent of the 850-bp promoter, mean  $\pm$  SE ( $n = 6$ ). **C** Luciferase activity as percent of the wild-type p470Luc, mean  $\pm$  SE ( $n = 12$ )

to binding sites for known transcription factors. Conserved element A contains the 11-bp sequence GCATAATTGAT which differs only at nucleotide 9 from the experimentally determined consensus for the transcription factor Pou6f1/Brn5 (Rhee et al. 1998), resulting in the high matrix similarity score of 0.918. Conserved element B has strong similarity to the YY1 consensus binding site, with a predictive score of 0.88. Direct repeats DR1 and DR2 match the consensus RE-1 site for binding the transcription factor REST/NRSF, with a lower score of 0.69. This is below the average matrix similarity for known functional RE-1 sites (Bruce et al. 2004; Mortazavi et al. 2006; Schoenherr et al. 1995). It is possible that the close juxtaposition of two weak sites could support functional interaction with the transcription factor. This analysis suggests that Pou6f1/Brn5, YY1, and REST/NRSF should be experimentally tested as candidates for regulation of *Scn8a* expression. Efforts to assess the occupancy of these sites using ChIP were unsuccessful due to the very high GC content of exon 1c and the surrounding genomic DNA (92% GC).

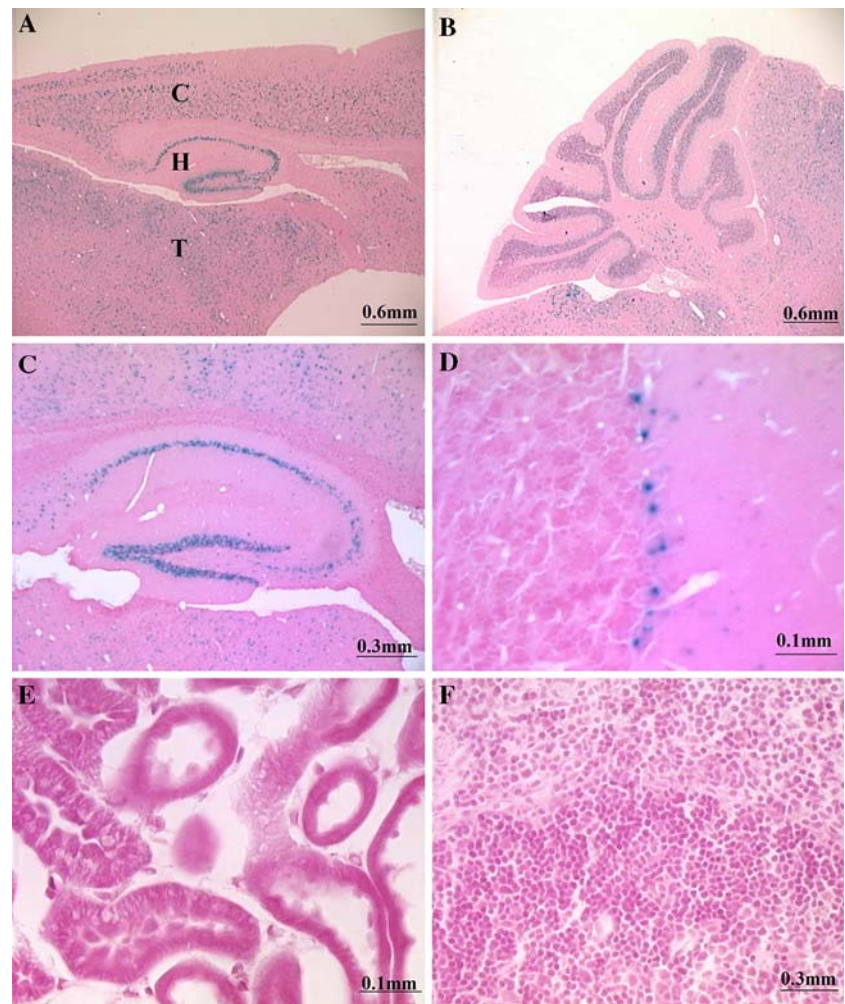
## Discussion

We previously compared the promoter regions of the human and mouse *SCN8A* genes only with the goal of identifying regulatory elements controlling gene transcription (Drews et al. 2005). The very high level of sequence conservation between human and mouse (93% in exon 1c) prevented the identification of small regulatory elements. In the current study, we extended the sequence comparison to chicken, fish, and opossum. Mammalian sequence could not be directly aligned with fish sequence, but the use of 5' RACE to locate the promoter region of the chicken gene enabled us to align the fish noncoding sequences. The unusually high sequence conservation of the 5' UTR of *SCN8A* (Table 1) is strongly indicative of conserved function. Elements within the 5' UTR have the potential to function in regulation of transcription or in post-transcriptional regulation of mRNA stability, localization, or translational efficiency. We tested the first possibility and demonstrated transcription-enhancing activity within the 5' noncoding exon of *SCN8A*.

Comparison of the organization of the 5' end of the vertebrate *SCN8A* genes provided an example of the general increase in gene size and complexity during evolution. More than half of the tenfold increase in length of the intronic sequence between the 5' noncoding exon and the first coding exon between fish and mammalian *SCN8A* is accounted for by the insertion of repetitive DNA. Another characteristic evolutionary trend is acquisition of multiple regulatory units controlling individual genes. Mammalian *SCN8A* appears to represent an intermediate stage in that process: In comparison with the one 5' noncoding exon in fish and chicken, the mammalian gene has four distinct 5' noncoding exons, but their expression patterns do not differ (Drews et al. 2005). The evolutionary comparison with chicken, fish, and marsupial genes indicates that mammalian exon 1c is descendent from the ancestral 5' noncoding sequence. The small proportion of 5'-RACE products containing exon 1c that we previously reported for human and mouse was an underestimate caused by the difficulty of amplification of this 93% GC fragment (Drews et al. 2005). When searching for transcripts containing the 5' UTRs of *Scn8a*, we found that 12 of the 13 mammalian ESTs in public databases contain exon 1c. Exon 1c was also present in 2/2 fish *Scn8a* ESTs, 4/4 *Xenopus Scn8a* ESTs, and 1/1 chicken EST (www.genome.ucsc.edu). Exon 1c thus appears to be the major first exon.

Within exon 1c four short elements exhibit a high proportion of evolutionarily invariant nucleotides (Fig. 1). Three lines of evidence indicate that exon 1c and its conserved elements are involved in transcriptional regulation. First, a 470-bp fragment containing exon 1c has strong promoter activity in transfected cells. Second, mutation of three of the four sites reduced the expression in transfected

**Fig. 4** Expression of p470LacZ in transgenic mice. The transgene structure is the same as Fig. 3 with LacZ reporter. Tissue sections were stained with Xgal to detect LacZ expression. **A** H, hippocampus; C, cortex; T, thalamus. **B** Cerebellum. **C** Hippocampus. **D** Cerebellar Purkinje cells. **E** Kidney. **F** spleen



cells. Third, the 470-bp fragment directs tissue-specific expression *in vivo* in transgenic mice. The close juxtaposition of four conserved elements within 120 bp suggests that they may function together to stabilize a protein complex during *SCN8A* transcription.

The conserved elements in exon 1c are related to binding sites for three transcription factors involved in neuronal gene regulation: Pou6f1, YY1, and REST/NSRF (Ballas and Mandel 2005; Coulson et al. 2005; Cui and Bulleit 1998; Gordon et al. 2006). *Pou6f1* has activity as a transcriptional enhancer (Molinari et al. 2004) and inhibitor (Wey et al. 1996) and is regulated by neuregulin in myelinating Schwann cells (Wu et al. 2001). *Scn8a* and *Pou6f1* are closely physically linked on mouse chromosome 15 and human chromosome 12q13. YY1 is a ubiquitously expressed member of the GLI-Kruppel family of zinc-finger transcription factors (Shi et al. 1997; Thomas and Seto 1999) that interacts with cofactors with histone modification activity (Thomas and Seto 1999). A YY1 element in reverse orientation, located at position +13 to +24 downstream of the transcription initiation site, is required for

transcription of the LINE-1 retrotransposon (Athaniar et al. 2004). The YY1 element in exon 1c is similarly located in reverse orientation downstream of the initiation site. The REST/NSRF transcription factor binds RE-1 elements and silences gene expression by a chromatin modification mechanism. Although the potential RE-1 sites in DR1 and DR2 do not exhibit a strong match to the consensus, the close juxtaposition of these two weak sites may result in functional binding.

We compared the sequence of exon 1c with other small promoter fragments that support brain-specific expression: a 107-bp fragment of *GnRH* (Kim et al. 2007), a 200-bp fragment of *Cyp19* (Nausch et al. 2007), and a 5-kb fragment of *BAl1-AP4* (Kim et al. 2004). None of the other promoters share sequences with exon 1c, indicating that there are alternative mechanisms for generating neuron-specific expression. The enhancers of *Sox2*, a transcription factor involved in neuronal development, are also distinct from exon 1c (Uchikawa et al. 2004).

We have used evolutionary approaches to localize a cluster of conserved sequences and performed initial

functional characterization. Further sequence comparisons with paralogous mammalian sodium channel genes, and with sodium channel genes from lower chordates, may elucidate the molecular basis for the divergence between brain-specific and muscle-specific sodium channels and identify elements that control the common expression patterns of the major mammalian neuronal channels *SCN1A*, *SCN2A*, *SCN3A*, and *SCN8A*.

Mutations in transcription factor binding sites contribute to the burden of genetic disease in man (Villard 2004). The biological importance of quantitative levels of *SCN8A* expression is clear from the haploinsufficiency of the human gene (Trudeau et al. 2006) and the phenotypic consequences of small differences in expression level of the mouse gene (Buchner et al. 2003). Altered expression of sodium channel genes is thought to play a role in the pathogenesis of human epilepsy. This report represents a first step toward understanding the molecular regulation of *SCN8A*.

**Acknowledgments** This work was supported by NIH Grant NS34509. For generation of transgenic mice, the authors acknowledge the contributions of Wanda Filipiak, Margaret Van Keuren, and the Transgenic Animal Model Core of the University of Michigan's Biomedical Research Core Facilities. Support for the Transgenic Animal Model Core was provided by The University of Michigan Center for Organogenesis, the Michigan Economic Development Corporation, and the Michigan Technology Tri-Corridor (Grant 085P1000815). The authors thank Michelle Trudeau for assistance with analysis of exon 1c alignments and completing the gap in the chicken genomic sequence, Andrew Lieberman for providing MN-1 cells and cell culture facilities, and Darcy Butts for assistance with the manuscript.

## References

- Athanikar JN, Badge RM, Moran JV (2004) A YY1-binding site is required for accurate human LINE-1 transcription initiation. *Nucleic Acids Res* 32:3846–3855
- Ballas N, Mandel G (2005) The many faces of REST oversee epigenetic programming of neuronal genes. *Curr Opin Neurobiol* 15:500–506
- Bruce AW, Donaldson IJ, Wood IC, Yerbury SA, Sadowski MI, et al. (2004) Genome-wide analysis of repressor element 1 silencing transcription factor/neuron-restrictive silencing factor (REST/NRSF) target genes. *Proc Natl Acad Sci USA* 101:10458–10463
- Buchner DA, Trudeau M, Meisler MH (2003) SCNMI, a putative RNA splicing factor that modifies disease severity in mice. *Science* 301:967–969
- Burgess DL, Kohrman DC, Galt J, Plummer NW, Jones JM, et al. (1995) Mutation of a new sodium channel gene, *Scn8a*, in the mouse mutant 'motor endplate disease.' *Nat Genet* 10:461–465
- Caldwell JH, Schaller KL, Lasher RS, Peles E, Levinson SR (2000) Sodium channel Na(v)1.6 is localized at nodes of ranvier, dendrites, and synapses. *Proc Natl Acad Sci USA* 97:5616–5620
- ICGS (2004) Sequence and comparative analyses of the chicken genome provide unique perspectives on vertebrate evolution. *Nature* 432:695–716
- Coulson JM (2005) Transcriptional regulation: cancer, neurons and the REST. *Curr Biol* 15:R665–668
- Cui H, Bulleit RF (1998) Expression of the POU transcription factor Brn-5 is an early event in the terminal differentiation of CNS neurons. *J Neurosci Res* 52:625–632
- Dodou E, Xu SM, Black BL (2003) *mef2c* is activated directly by myogenic basic helix-loop-helix proteins during skeletal muscle development in vivo. *Mech Dev* 120:1021–1032
- Drews VL, Lieberman AP, Meisler MH (2005) Multiple transcripts of sodium channel SCN8A (Na(V)1.6) with alternative 5'- and 3'-untranslated regions and initial characterization of the SCN8A promoter. *Genomics* 85:245–257
- Ghanem N, Jarinova O, Amores A, Long Q, Hatch G, et al. (2003) Regulatory roles of conserved intergenic domains in vertebrate *Dlx* bigene clusters. *Genome Res* 13:533–543
- Gordon S, Akopyan G, Garban H, Bonavida B (2006) Transcription factor YY1: structure, function, and therapeutic implications in cancer biology. *Oncogene* 25:1125–1142
- Kim HH, Wolfe A, Cohen RN, Eames SC, Johnson AL, et al. (2007) In vivo identification of a 107-base pair promoter element mediating neuron-specific expression of mouse gonadotropin-releasing hormone. *Mol Endocrinol* 21:457–471
- Kim MY, Ahn KY, Lee SM, Koh JT, Chun BJ, et al. (2004) The promoter of brain-specific angiogenesis inhibitor 1-associated protein 4 drives developmentally targeted transgene expression mainly in adult cerebral cortex and hippocampus. *FEBS Lett* 566:87–94
- Levin SI, Khaliq ZM, Aman TK, Grieco TM, Kearney JA, et al. (2006) Impaired motor function in mice with cell-specific knockout of sodium channel *Scn8a* (NaV1.6) in cerebellar purkinje neurons and granule cells. *J Neurophysiol* 96:785–793
- Lopreato GF, Lu Y, Southwell A, Atkinson NS, Hillis DM, et al. (2001) Evolution and divergence of sodium channel genes in vertebrates. *Proc Natl Acad Sci USA* 98:7588–7592
- Makalowski W, Boguski MS (1998) Evolutionary parameters of the transcribed mammalian genome: an analysis of 2,820 orthologous rodent and human sequences. *Proc Natl Acad Sci USA* 95:9407–9412
- Meisler MH, Plummer NW, Burgess DL, Buchner DA, Sprunger LK (2004) Allelic mutations of the sodium channel SCN8A reveal multiple cellular and physiological functions. *Genetica* 122:37–45
- Molinari S, Relaix F, Lemonnier M, Kirschbaum B, Schafer B, et al. (2004) A novel complex regulates cardiac actin gene expression through interaction of Emb, a class VI POU domain protein, MEF2D, and the histone transacetylase p300. *Mol Cell Biol* 24:2944–2957
- Mortazavi A, Thompson EC, Garcia ST, Myers RM, Wold B (2006) Comparative genomics modeling of the NRSF/REST repressor network: from single conserved sites to genome-wide repertoire. *Genome Res* 16:1208–1221
- Nausch N, Manteuffel G, Vanselow J (2007) 0.2-kb promoter sequence of the murine *Cyp19* gene target beta-galactosidase expression to specific brain areas of transgenic mice. *J Steroid Biochem Mol Biol* 103:119–128
- Nobrega MA, Ovcharenko I, Afzal V, Rubin EM (2003) Scanning human gene deserts for long-range enhancers. *Science* 302:413
- Novak AE, Jost MC, Lu Y, Taylor AD, Zakon HH, et al. (2006) Gene duplications and evolution of vertebrate voltage-gated sodium channels. *J Mol Evol* 63:208–221
- Okamoto K, Wakamiya M, Noji S, Koyama E, Taniguchi S, et al. (1993) A novel class of murine POU gene predominantly expressed in central nervous system. *J Biol Chem* 268:7449–7457
- Rhee JM, Gruber CA, Brodie TB, Trieu M, Turner EE (1998) Highly cooperative homodimerization is a conserved property of neural POU proteins. *J Biol Chem* 273:34196–34205



- Schaller KL, Krzemien DM, Yarowsky PJ, Krueger BK, Caldwell JH (1995) A novel, abundant sodium channel expressed in neurons and glia. *J Neurosci* 15:3231–3242
- Schoenherr CJ, Anderson DJ (1995) The neuron-restrictive silencer factor (NRSF): a coordinate repressor of multiple neuron-specific genes. *Science* 267:1360–1363
- Shi Y, Lee JS, Galvin KM (1997) Everything you have ever wanted to know about Yin Yang 1. *Biochim Biophys Acta* 1332:F49–66
- Spitz F, Gonzalez F, Duboule D (2003) A global control region defines a chromosomal regulatory landscape containing the HoxD cluster. *Cell* 113:405–417
- Thomas MJ, Seto E (1999) Unlocking the mechanisms of transcription factor YY1: are chromatin modifying enzymes the key? *Gene* 236:197–208
- Trudeau MM, Dalton JC, Day JW, Ranum LP, Meisler MH (2006) Heterozygosity for a protein truncation mutation of sodium channel SCN8A in a patient with cerebellar atrophy, ataxia, and mental retardation. *J Med Genet* 43:527–530
- Uchikawa M, Takemoto T, Kamachi Y, Kondoh H (2004) Efficient identification of regulatory sequences in the chicken genome by a powerful combination of embryo electroporation and genome comparison. *Mech Dev* 121:1145–1158
- Villard J (2004) Transcription regulation and human diseases. *Swiss Med Wkly* 134:571–579
- Wey E, Schafer BW (1996) Identification of novel DNA binding sites recognized by the transcription factor mPOU (POU6F1). *Biochem Biophys Res Commun* 220:274–279
- Wu R, Jurek M, Sundarababu S, Weinstein DE (2001) The POU gene Brn-5 is induced by neuregulin and is restricted to myelinating Schwann cells. *Mol Cell Neurosci* 17:683–695
- Zerucha T, Stuhmer T, Hatch G, Park BK, Long Q, et al. (2000) A highly conserved enhancer in the *Dlx5/Dlx6* intergenic region is the site of cross-regulatory interactions between *Dlx* genes in the embryonic forebrain. *J Neurosci* 20:709–721