Available online at www.sciencedirect.com
SCIENCE DIRECT
Biochemical and Biophysical Research Communications 344 (2006) 834–844
BBRC
www.elsevier.com/locate/ybbrc

Molecular cloning and characterization of human *Castor*, a novel human gene upregulated during cell differentiation ☆

Zhihui Liu [1], Xuezhong Yang [1], Fei Tan, Kathleen Cullion, Carol J. Thiele *

Cell and Molecular Biology Section, Pediatric Oncology Branch, National Cancer Institute, National Institutes of Health, Bethesda, MD 20892, USA

Received 16 March 2006
Available online 19 April 2006

Abstract

Castor is a zinc finger transcription factor that controls cell fate within neuroblast cell lineages in *Drosophila melanogaster*. Here, we describe the cloning and characterization of a human *castor* gene (*CASZ1*) that is structurally homologous to *Drosophila castor*. We find the expression of *castor* gene is increased when cells of neural origin as well as mesenchymal origin are induced to differentiation. *CASZ1* is expressed in a number of normal tissues and exists in at least two mRNA species of 4.4 and 8.0 kb. They are named *hCasz5* and *hCasz11* because the predicted proteins have 5 and 11 zinc fingers, respectively. Deletion analysis of the proximal 5′-flanking sequences delineates sequences sufficient to drive transcription in cells of neural and non-neural origin. Both *hCasz5* and *hCasz11* localize predominantly in the nucleus, consistent with their role as Zn-finger containing transcription factor. *CASZ1* is expressed in a number of human tumors and localizes to a chromosomal region frequently lost in tumors of neuroectodermal origin.
Published by Elsevier Inc.

Keywords: *Castor*; *CASZ1*; Transcription factor; Cell differentiation; Zinc finger; Neuroblastoma; Promoter

The *Drosophila castor* gene (*dCas*) encodes a zinc finger protein that is expressed in a specific subset of neuroglioblasts in the ventral nerve chord and the procephalic region of *Drosophila* during embryogenesis. *DCas* is required postembryonically for correct axon pathfinding of the central complex and mushroom body neurons. Loss of *dCas* function results in precise alterations in central nervous system gene expression, defects in axonogenesis, and embryonic lethality [1–7]. *DCas* is expressed exclusively at a late stage of neuroblast development that precedes the cessation of proliferation and the beginning of neuronal differentiation. In *Drosophila* it plays a central role in controlling cell fates within neuroblast cell lineages. The stage during which castor is expressed during *Drosophila* development is analogous to later stages of human embryonic neural organogenesis [8] which implies that its homolog-human *castor* (*CASZ1*) may have the same role as *dCas* in controlling cell fates within neuroblast cell lineages.

As a pre-requisite to studies to evaluate whether *CASZ1* plays a role in neural fate determination in humans, we cloned and characterized the human *CASZ1* gene. In this study, we report the cloning of the human *CASZ1* gene and the 5′ proximal flanking region of the gene that drives transcription from a reporter gene. Our analysis reveals in addition to cloning the putative CASZ1 transcript encoding five Zn fingers, we cloned an additional novel CASZ1 isoform which contains 11 Zn fingers. A high degree of homology is found among the *human* CASZ1 and hypothetical proteins from several other species including *Drosophila melanogaster*, *Mus musculus*, *Rattus norvegicus*, *Canis familiaris*, and *Tetraodon nigroviridis castor*. Such a high degree of evolutionary conservation is usually associated with genes with functional importance.

☆ *Abbreviations:* NB, neuroblastoma; *CASZ1*, *Homo sapiens castor*; *mCas*, *Mus musculus castor*; *rCas*, *Rattus norvegicus castor*; *cCas*, *Canis familiaris castor*; *tCas*, *Tetraodon nigroviridis castor*; ORF, open-reading frame; EST, express sequence tag; ZnF, zinc finger; AA, amino acid; NLS, nuclear localization sequence.

* Corresponding author. Fax: +1 301 451 7052.
E-mail address: ct47a@nih.gov (C.J. Thiele).

[1] These authors contributed equally and are co-first authors.

0006-291X/$ - see front matter Published by Elsevier Inc.
doi:10.1016/j.bbrc.2006.03.207

Materials and methods

Cell culture. The following cell lines were used in this study: Neuroblastoma—SMS-KCNR, SK-N-AS, SK-N-BE2, SK-N-DZ, SK-N-LE, NBLS, and SY5Y; Glioblastoma—T98G; Retinoblastoma—W27, Y79; Rhabdomyosarcoma—A204, RD, and RH30; Ewing's sarcoma—TC71, TC106, and TC268; Cervical Adenocarcinoma—HeLa; Myoblasts—C2C12; Fibroblast—COS7, NIH3T3. Mouse C2C12 myoblasts were maintained in DMEM supplemented with 10% fetal bovine serum, 50 U/ml penicillian, and 50 μg/ml streptomycin. The other cells were cultured in RPMI 1640 (Mediatech, Herndon, VA) containing 10% fetal bovine serum, 2 mM glutamine, and antibiotics as previously described [9]. To induce cell differentiation, C2C12 were plated into 6-well plate at 500,000/well, 24 h later when cells were 80% confluent, the medium was switched to the differentiation medium containing DMEM and 2% horse serum, which was changed every other day; KCNR cells were plated into 6-well plate at 400,000/well, 24 h later, the medium was replaced with RPMI 1640 containing 1 mM dibutyryl cyclic AMP. Cells were transfected with plasmids or small interference RNA by using Lipofectamine™ 2000 (Invitrogen, Carlsbad, CA) following the manufacturer's protocol.

Northern blot analysis. Total RNA from cell lines and different parts of postnatal day 6 mice brain was isolated with RNeasy kit (Qiagen, Valencia, CA) following the manufacturer's protocol. Fifteen micrograms of total RNA was run on 1% formaldehyde gel and blotted to Nytran filter. For Human Adult Normal Tissue Total RNA Northern Blot I, II (Clontech, Mountain View, CA), just following the manufacturer's protocol. The blot was then hybridized with indicated probes isolated from plasmids containing *CASZ1* gene fragment from 3477 to 5087 bp. The probes were labeled with ^{32}P by Rediprime™ II (Amersham Biosciences, Buchinghamshire, England).

Real-time PCR. Two pairs of primer were designed for *CASZ1* mRNA detection, the primer pair 1 amplified 130 bp of *hcasz5* 4406 bp transcript from 1619 to 1748; the primer pair 2 amplified 149 bp of *hcasz11* 7946 transcript from 4358 to 4485 and contains no binding site in *hcasz5*. A specific primer pair that amplified 120 bp of the human housekeeping gene β-actin was used as control. The primer for mouse *castor* amplified 134 bp of mouse *castor* transcript (GenBank Accession No. BC_075673) from 761 to 894. A specific primer pair that amplified 103 bp of the mouse house keeping gene β-actin was used as control. Quantitative real-time PCR was performed on ABI Prism 7000 (PE Applied Biosystems) by using Platinum® SYBR® Green qPCR SuperMix-UDG (Invitrogen, Carlsbad, CA) following manufacturer's instructions.

RNAi. We designed and purchased *hcasz5* siRNA oligonucleotides from Qiagen, the target sequence was 5′-AACGGACTGCCCACAGA TAAA-3′ and located at 1662–1682 of *hcasz5* 4406 bp transcript (GenBank Accession No. AK092289). Non-silencing siRNA (Qiagen Cat. No. 1022076100) was used as negative control. One hundred micromolar of siRNA was transfected into SK-N-As cells by using Lipofectamine™ 2000. Forty-eight hours later, cells were harvested for RNA extraction. Total RNAs were extracted by using RNAeasy kit (Qiagen, Valencia, CA) according to the manufacturer' s instructions. Two micrograms of total RNA was reverse-transcribed using superscript III reverse transcriptase (Invitrogen, Carlsbad, CA).

Cloning of the 5′-flanking region of CASZ1 gene and activity assay. The transcription start site was determined by 5′ RACE. Total RNA was extracted from human muscle tissue and an adapter molecule was ligated to the 5′ end of integrate mRNA with 5′ cap using the GeneRacer kit (Invitrogen, CA). The cDNA was synthesized from the ligated RNA using superscript II. A gene-specific primer (5GA2: CATCATAGCCAAACAT GCCGAG) and the adapter primer were combined to amplify the 5′ end of the cDNA, and a nested PCR was performed to further amplify the specific end of the target gene using a nested gene-specific primer (5GA1: CGTCCGAATCCTTCTCCTCC) and a nested adapter primer. The amplified end of the cDNA was cloned, sequenced, and confirmed to be consistent with the most 5′ sequence of AK092289 (less 2 bp). This represents the most 5′ sequence of all of sequences representing *CASZ1* gene as documented in GenBank. Based on this result, the different sequences from the promoter region of the *CASZ1* gene consisting of different lengths from 3298 to 18 bp of 5′-flanking sequences upstream of the exon 1 and containing the first 62 bp of the exon 1 were cloned into the promoter less luciferase reporter vector pGL2-basic (Promega). The selected region(s) was amplified by PCR using normal human tissue DNA as template and all PCR products were verified by DNA sequence analysis. Individual reporter plasmids and pCMV-β-gal were co-transfected into AS cells (1×10^6), and the luciferase and β-galactosidase activity assays were performed as described previously [10,11].

CASZ1 gene cloning and sequencing. The partial cDNA of CASZ1 was used to screen commercial cDNA libraries to obtain the *hcasz5* clone (GenBank Accession No. AK092289, clone ID NTONG2005363) from National Institute of Technology and Evaluation (NITE, Japan). The *hcasz11* 3′ terminal fragment clone was obtained from Origene (clone BC1272_C01). Since there is no overlap between clone NTONG2005363 and clone BC1272_C01, a specific primer pair was created to amplify the middle region of the *hcasz11* cDNA resulting in a 1.6 kb PCR product that overlaps with the 3′ terminal of *castor* fragment in NTONG2005363 and the 5′ terminal of *castor* fragment in BC1272_C01. The primers chosen were 5′-CTTCCACTTCCGGACAGAGGGAG-3′ and 5′-GAGTCCAT CTGCGACATGCC-3′. A retina cDNA library was used as template for RT-PCR. PCR fragments were gel purified using Gel Extraction Kit (Qiagen Inc., Mississauga, ON, Canada) and cloned into pGEM T vector (Promega, Madison, WI). Sequencing of the clone was performed by DNA Sequencer according to the manufacturer's protocol. The insert from the correct clone was cut out by *Sma*I/*Bsp*EI and ligated into NTONG2005363 that had been treated with *Not*I/DNA polymerase I large fragment and *Bsp*EI, in order to generate a clone with elongated *hcasz5*. This clone was further digested by *Eco*RI/*Bcl*I and the insert was cloned into the *Eco*RI/*Bcl*I restricted BC1272_C01 vector. The *CASZ1* 7746 bp cDNA was cloned into the vector pCMV6-XL5 and named pCMV6-XL5 *hcasz11*.

Construction of CASZ1 FLAG-tagged fusion protein. The entire coding region of the novel cDNA was fused inframe 3′ to a FLAG tag in the vector pCMV-Tag 2A (Stratagene, La Jolla, CA). To prepare the FLAG-*hcasz5* and FLAG-*hcasz11* expression vectors, the ORF fragment of *CASZ1* 5′ terminal was amplified by PCR using appropriate primers (forward Primer, 5′-CGCGGATCCAGATCTTGGAACAGCTGAG-3′; reverse primer, 5′-CCGGAATTCGACTCGGGGTCCGCTTCTCCAC CACC-3′) containing *Bam*HI and *Eco*RI/*Ahd*I sites at the 5′- and 3′-ends, respectively. The PCR products were inserted into *Bam*HI/*Eco*RI sites of the pCMV-Tag 2A (Stratagene). *Hcasz5* containing clone NTONG2005363 was treated by *Sbf*I and *Nhe*I/DNA polymerase I large fragment, then we inserted it into pCMV-Tag 2A with *CASZ1* 5′ terminal fragment that has been treated with *Sbf*I and *Ahd*I/ DNA polymerase I large fragment, we obtained pCMV-Tag 2A *hcasz5* with full *hcasz5* ORF. *Hcasz11* was cut off from pCMV6-XL5 *hcasz11* by *Sbf*I and *Ahd*I, then inserted into *Sbf*I/*Ahd*I sites of pCMV-Tag 2A with *CASZ1* 5′ terminal fragment and pCMV-Tag 2A *hcasz11* with full *hcasz11* ORF was obtained. The vector-insert junction was verified by nucleotide sequencing.

Western blot analysis. Nuclear and cytoplasmic proteins were extracted using the nuclear extraction kit (IMGENEX, San Diego, CA) following the manufacturer's protocol. Total cell protein and Western blot analysis was performed as described previously [10]. Briefly, total protein, nuclear and cytoplasmic proteins (20 μg) were analyzed by 4–12% SDS–PAGE and transferred to nitrocellulose. As a primary antibody, the anti-FLAG M2 monoclonal antibody was diluted, (1:1000) incubated for 1 h and then was detected with horseradish peroxidase-conjugated goat anti-mouse IgG (1:2000 dilution).

Immunocytochemistry. SK-N-AS cells transfected with FLAG-*hcasz5* and FLAG-*hcasz11* were grown on coverslips for 48 h, fixed in 4% paraformaldehyde in 1× PBS for 15 min at room temperature, and permeabilized with 0.25% Triton X-100 for 5 min. Anti-FLAG M2 monoclonal antibody (Sigma–Aldrich) was used as primary antibody (1:200 dilution) and a rhodamine-labeled goat anti-mouse IgG (Santa cruz) was used as secondary antibody (1:100 dilution). Controls for the immunostaining were the cells transfected with empty vector.

In silico analysis. We employed blastn program http://www.ncbi.nlm.nih.gov/BLAST/ [12] to search GenBank and EST databases to obtain the full-length cDNA containing start-, stop-codon, poly(A) signal and a complete open-reading frame (ORF) of *CASZ1* gene. *CASZ1* DNA and Genomic DNA alignment was performed by BLAT program http://genome.ucsc.edu/. Transcription factors and their binding sites were predicted by employing TRANSFAC database http://www.gene-regulation.com/index.html [13] and MAPPER program http://mapper.chip.org/ that combines TRANSFAC and JASPAR data [14]. A BLAST search (www.ncbi.nlm.nih.gov/BLAST) [12] for related proteins was performed. Comparisons of related amino acid sequences and the display of the resulting alignments were performed using DIALIGN 2.2.1 program http://bioweb.pasteur.fr/seqanal/interfaces/dialign2-simple.html [15]. Protein motifs were searched using software available on these web sites: http://expasy.ch/prosite, http://smart.embl-heidelberg.de [16,17] and http://myhits.isb-sib.ch/cgi-bin/motif_scan (Swiss Institute of Bioinformatics).

GenBank Accession Number. The cDNA sequences for the *CASZ1* gene family have been deposited in GenBank and assigned Accession No. DQ217660.

Results

To determine the size and relative level of *CASZ1* transcripts in human tissues, Northern blot analyses of RNA from normal human tissues and cell lines from tumors of neuroectodermal origin were hybridized with PCR fragments amplified from plasmids containing *CASZ1* fragments as described in Materials and methods. Two predominant RNA species of approximately 8.0 and 4.4 kb hybridize on the Northern blots with ^{32}P-oligos for *CASZ1* (Fig. 1). The *castor* gene is highly expressed in human heart, lung, skeletal muscle, pancreas, testis, small intestine, and stomach, but it is not detectable in the adult brain (Fig. 1A). However, using real-time PCR, *castor* gene expression is detected in RNA from postnatal day 6 mouse brain with levels of expression being relatively higher in brain stem and the thalamencephalon than in cerebellum (Fig. 1B). In neuroblastoma (NB) cell lines, tumors of the peripheral nervous system, both the 8.0 and 4.4 kb *CASZ1* mRNAs are expressed in SKNLE, SKNDZ, BE2, SY5Y, NBLS, AS, and KCNR although the levels varied over 10-fold among these tumors. The two *CASZ1* transcripts are detected in other tumors of neuroectodermal origin Ewing's sarcoma (TC71) and a retinoblastoma (Y79) as well as a tumor of mesodermal origin, embryonal rhabdomyosarcoma (RD), and monkey kidney cells (COS) while in HeLa, NIH3T3, and an alveolar rhabdomyosarcoma (RH30) the expression of *CASZ1* is barely detectable (Fig. 1C). By real-time PCR, *CASZ1* mRNA is detected invariably detected in other cell lines of neuroectodermal origin cell lines including malignant rhabdoid tumor (A204), glioblastoma (T98G), and EWS (TC268) as well as a Wilm's tumor cell line that is derived from kidney cells (W27). The only cell line tested that lacked castor expression was a neuroectodermal cell line TC106 (Fig. 1D).

As *castor* is a developmentally regulated gene in *Drosophila*, we utilized two in vitro models of cell differentiation to assess whether CASZ1 expression was regulated during differentiation. Since *castor* is highly expressed in

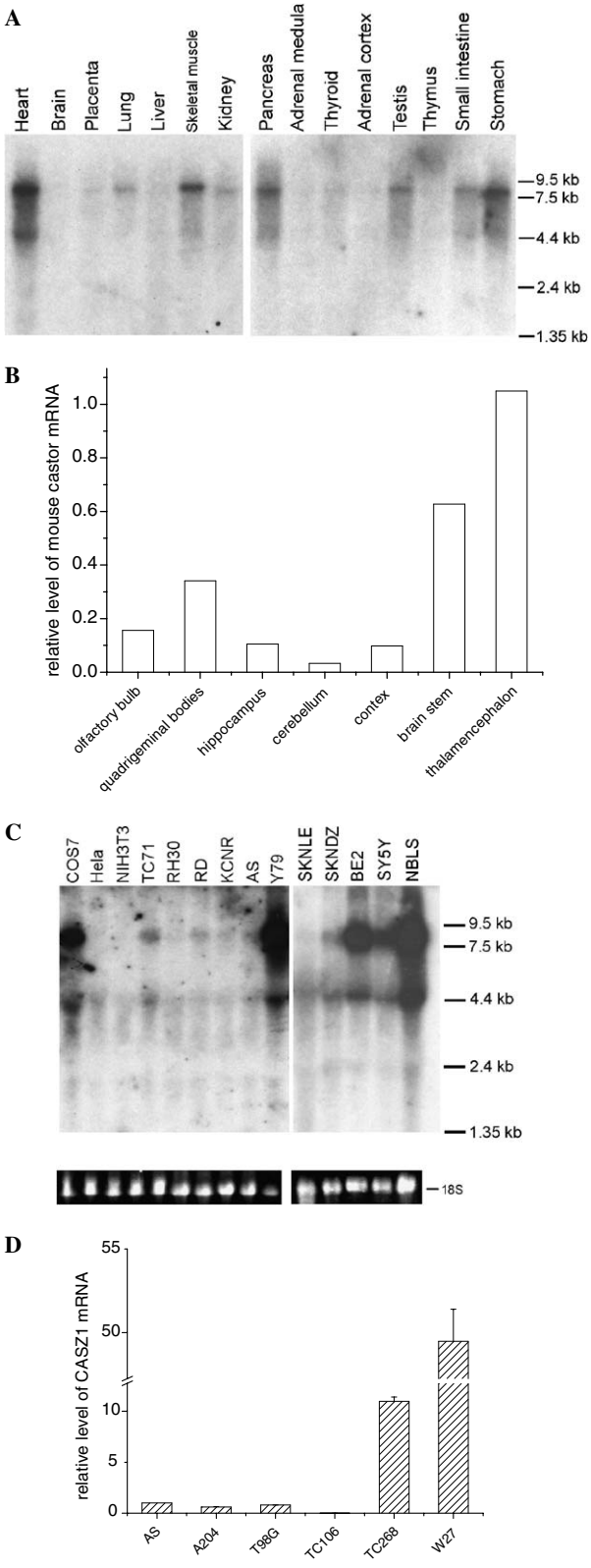

Fig. 1. Expression profile of *CASZ1* in various human adult tissues (A) and cell lines (C) detected by Northern blot. Two transcripts were detected, the longer one is about 8.0 kb and the short one is about 4.5 kb. Relative level of *castor* mRNA in different parts of postnatal day 6 mice brain and some other cell lines was detected by real-time PCR (B,D).

adult muscle, we examined its expression in a mouse model of muscle differentiation. Murine myoblasts (C2C12) can be induced to differentiate and form multi-nucleated myotube when cultured in 2% horse serum [18]. Mouse *castor* mRNA levels were dramatically upregulated during myotube formation in C2C12 cells. Four days after being shifted to differentiation medium, there was a 40-fold increase in *castor* expression and by 8 days the increase in *castor* expression was 135-fold greater than the levels in control cells at day 0 (Fig. 2A). The increase in *castor* expression coincided with the formation of myotube at day 4 and the levels further increased in myotube at day 8 (Fig. 2B). Although in our Northern blots, *CASZ1* expression was barely detected in adult brain, we examined its regulation in neural cells using human neuroblastoma cells that can be induced to extend neurite-like processes when treated with agents that raise intracellular cAMP levels such as dibutyryl cAMP [19]. In this model, *CASZ1* mRNA was significantly upregulated in neuroblastoma cells with levels increasing almost 2.4-fold after 24 h and

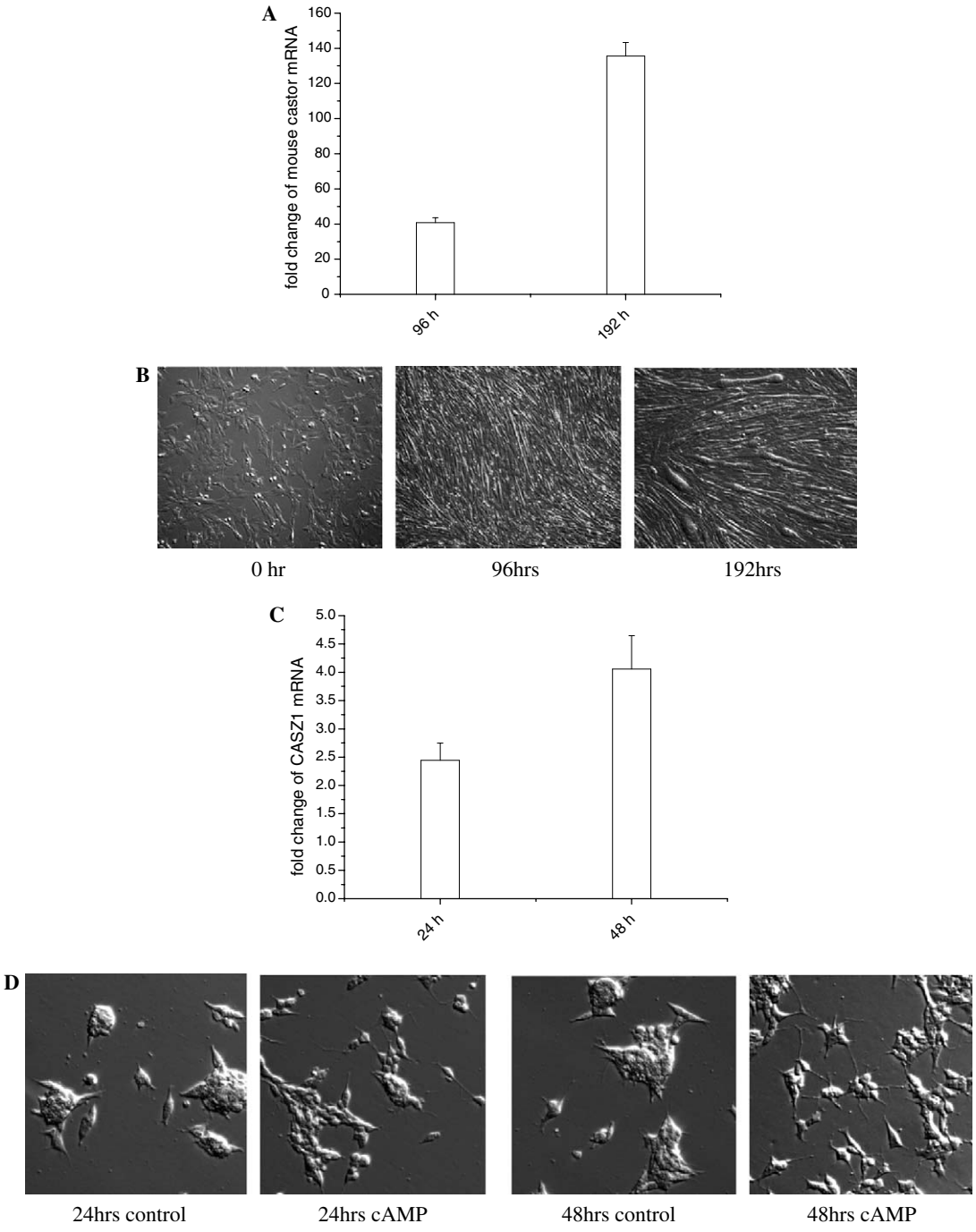

Fig. 2. The relationship between *castor* gene expression and cell differentiation detected by real-time PCR. The *castor* gene expression was increased in myoblast cells (C2C12) during formation of myotubes (A,B) and in neuroblastoma cells (KCNR) during neurite extension (C,D).

4-fold after 48 h of dibutyryl cyclic AMP treatment compared to untreated control cells (Fig. 2C). Increases in *CASZ1* mRNA were associated with the extension of neuritis, a hallmark of neuronal differentiation (Fig. 2D).

By using blastn program, the short transcript of *CASZ1* was found in the GenBank and defined as castor homolog 1, zinc finger (Gene name: CASZ1; GeneID: 54897; GenBank Accession No. AK092289). Consistent with our Northern blot analysis the cDNA is 4406 bp long and encodes a putative protein of 1166 amino acids with 5 TFIIIA class C2H2 zinc finger (ZnF) motifs and highly conserved *para*-ZnF sequences that are unique to *Drosophila* castor (Fig. 3). We named this isoform *hcasz5*. The CDS is complete since there is an inframe stop in the 5′ UTR 39 bp before the start codon (ATG start from 321 bp). Our Northern blot analysis indicates the existence of an additional mRNA species with homology to *castor*. By using the AK092289 mRNA sequence as template, we performed in silico analysis utilizing human mRNAs and ESTs from GenBank to discover the longer *castor* mRNA that results from alternative splicing. We acquired a longer cDNA candidate that contains a putative start-, stop-codon and poly(A) signal. The mRNAs and ESTs that overlapped each other sequentially were as follows: GenBank Accession Nos. AK130996, AK022285, AK022482, AK021990, AK092289, AX747454, AK000328, BC004410, BC051883, BU729694, BP239918, BM700915, R85589, BQ188241, BQ423979, DN831131, AA449398, R50443, AW812234, AW812251, BM048706, BM793269, BM795595, AA021322, AA018672, AA410866, BQ083003, AA015880, AA059087, CN427249, BG984806, BM563826, BE003506, BM800346, and CN427248. In order to simplify the cloning of the longer *CASZ1* mRNA, we screened commercial cDNA libraries and found the 3′ terminal fragment of a *CASZ1* transcript that overlapped with the above ESTs from Origene Company. The clone, BC1272-C01, consists of a 3592 bp cDNA. By combining the above mRNA, ESTs, and cDNA clone, a longer *CASZ1* cDNA is obtained. This cDNA comprises 7946 bp and encodes a 1759 amino acid protein with 11 TFIIIA class C2H2 zinc finger (ZnF) motifs and a highly conserved *para*-ZnF sequence, so we named this sequence *hcasz11*.

Alignment of the *hcasz5* transcript (GenBank Accession No. AK092289) and *hcasz11* transcript sequences to genomic DNA revealed sequences with 16 and 21 exons, respectively (Fig. 4A). By comparing the *hcasz5* with *hcasz11* transcript sequences, we found that in exon 16 of *hcasz5* an additional splice site can provide a 5′ splice donor. This 5′ splice donor in *hcasz5* is located just 1 bp before the stop codon of *hcasz5* in exon 16 and by splicing it with the 3′ splice acceptor sequences in exon 17, *hcasz11* is produced (Fig. 4A).

The premessenger of *hcasz5* has 16 exons and covers 149 kb on chromosome 1 (p36.22) from 10,641,535 to 10,790,973 using the complementary chain as template for transcription. The *hcasz5* 5′ UTR contains 320 bp and the 3′ UTR contains about 585 bp followed by poly(A) site. The start codon is located at exon 1 and the stop codon located at exon 16, the open-reading frame (ORF) is from 321 to 3818 bp. Within the 3′ UTR there are two putative consensus polyadenylation signals (ATAAA, nt: 4380–4384; 4389–4393) and one RNA instability motif (ATTTA, nt: 4401–4405). The premessenger of *hcasz11* has 21 exons and covers 160 kb on chromosome 1 (p36.22) from 10,630,927 to 10,790,973 using the complementary chain as a template for transcription. *Hcasz11* 5′ UTR contains about 320 bp and 3′ UTR contains 2364 bp followed by poly(A). The start codon located at exon 1 and the stop codon located at exon 21, the open-reading frame (ORF) was from 321 to 5597 bp (Figs. 4A and B). Within the *hcasz11* 3′ UTR there are two putative consensus polyadenylation signals (ATAAA, nt: 7618–7622; 7919–7923) and four RNA instability motifs (ATTTA, nt: 6661–6665; 7824–7828; 7841–7845; 7859–7863).

Hcasz5 4406 bp mRNA transcript encodes a protein of 1166 amino acids (AA) (124.7 kDa, p*I* 8.4). Based on the SMART tool analysis result, we found *hcasz5* has five zinc fingers and belongs to a classic *Xenopus laevis* transcription factor TFIIIA zinc finger-like protein. However, we found that there is a highly conserved *para*-ZnF sequence located upstream of each classic zinc finger, a feature that is identical to the zinc fingers in *Drosophila castor* (Fig. 3). The zinc fingers of *hcasz5* indicate that it has zinc ion-binding (GO: 0008270) and nucleic acid-binding (G: 0003676) molecular functions. The ScanProsite tool scanning results from Prosite database identify a Nuclear Locating Sequence (NLS) located at: 232–248 indicating that *hcasz5*

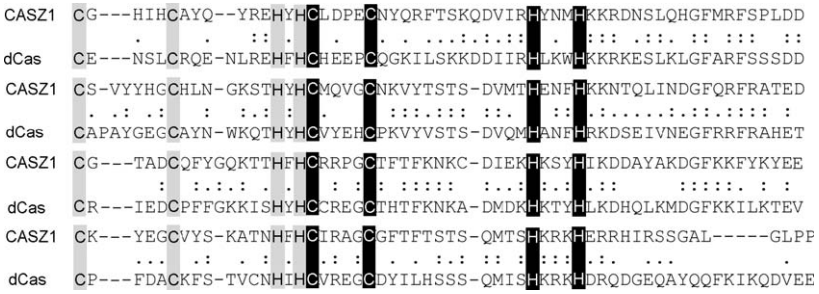

Fig. 3. Sequence comparison of C2-H2C2-H2 repeats from *CASZ1* and *Drosophila castor (dCas)*. First four zinc fingers of *CASZ1* were included. The conserved TFIIIA class C2H2 zinc finger (ZnF) motifs C and H residues are dark highlighted and the novel *para*-ZnF motifs C and H residues are gray highlighted. For optimal alignment, gaps have been introduced as indicated by horizontal bars.

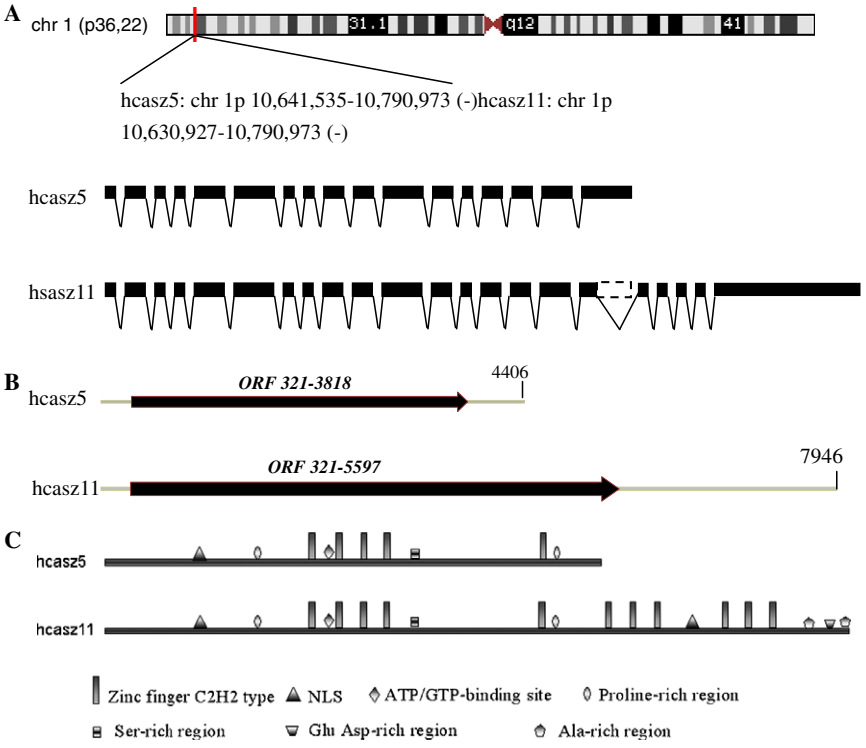

Fig. 4. Genomic structure, cDNA information, and the predicted protein motifs of *CASZ1*. (A) Representation of genomic structure for *CASZ1*. The *CASZ1* gene is located in chr1p 36.22. The gene encodes two transcripts, *hcasz5* and *hcasz11*. Exon 16 is differently spliced to produce these two transcripts. (B) The assembled *CASZ1* comprises a 5′ UTR region/3′ UTR region, start/stop-codon, and poly(A) signal. Within the 3′ UTR there are two putative consensus polyadenylation signals (ATAAA) for both *hcasz5* (nt: 4380–4384; 4389–4393) and *hcasz11* (nt: 7618–7622; 7919–7923). There is one RNA instability motif (ATTTA) for *hcasz5* (nt: 4401–4405) and there are four for *hcasz11* (nt: 6661–6665; 7824–7828; 7841–7845; 7859–7863). (C) Both *hcasz5* and *hcasz11* proteins have the motif such as zinc finger, NLS, ATP-binding site, Ser-rich region, and proline-rich domain, *hcasz11* also has Glu Asp-rich region and Ala-rich region.

localizes to the nucleus (GO: 0005634). According to the above bioinformation and the function of its homolog *Drosophila castor*, *hcasz5* should be a transcription factor. The motif scan also identifies an ATP/GTP-binding site motif A (Prosite: ps00017) at 541–548. Moreover, the presence of a Ser-rich region located at 720–750 and two Proline-rich regions located at 384–417 and 1080–1143 indicates potential protein–protein interactions. The sketch map of motifs identified in *hcasz5* is shown in Fig. 4C.

Hcasz11 cDNA encodes a protein of 1759 AA (190.0 kDa, p*I* 6.64) with the first 1166 AA identical to *hcasz5*. Analysis of the putative *hcasz11* protein sequence indicates that in addition to containing all of the information included in *hcasz5*, *hcasz11* (Fig. 4C) has an additional NLS located at 1401–1418, a Glu Asp-rich region located at 1672–1729, and two Ala-rich regions located at 1635–1670 and 1726–1756. Based on SMART analysis, *hcasz11* has another 6 TFIIIA class C2H2 zinc finger motifs that also contain the highly conserved *para*-ZnF sequence as found in the first five zinc fingers in *hcasz5*.

Aside from the homology between *CASZ1* and *dCas*, the *CASZ1* genes showed very strong homologies at both nucleotide and amino acid sequences to the *M. musculus*, *R. norvegicus*, *Canis familiaris*, and *tetraodon nigroviridis castor* genes (*mCas*, *rCas*, *cCas*, and *tCas*) (Table 1). *mCas*, *rCas*, and *cCas* have the similar length to *hcasz5* and the identity is over 85% both at CDS and protein level. The CDS of *tCas* has 5298 bp DNA and encodes a 1765 AA protein (NCB Accession CAG07550), which is as long as *hcasz11*. *TCas* and *hcasz11* are 58% homologous for CDS nucleotide identity and 63% for AA identity. The DIALIGN 2.2.1 program [15] was adopted for the protein alignment and the results are shown in Supplement Fig. 1. These results indicate that *castor* is evolutionarily conserved in divergent species range.

Table 1
Percent identities of CASZ1 cDNA and protein among several species

	hcasz5/ dCas	*hcasz5/ mCas*	*hcasz5/ rCas*	*hcasz5/ cCas*	*hcasz11/ tCas*
CDS identity (%)	—	87	87	89	58
Protein identity (%)	35	90	85	90	63

Note: mCas: Mus musculus castor (GenBank Accession No. XM_112612); *rCas: Rattus norvegicus castor* (GenBank Accession No. XM_243038); *cCas: Canis familiaris castor* (GenBank Accession No. XM_544573); *tCas*: *Tetraodon nigroviridis castor* (GenBank Accession No. CAAE01014991); the *hcasz5* and *dCas* CDS identity is undetectable.

Based on the expression studies by Northern blot analysis and in silico analysis, at least two splicing transcripts of *CASZ1* exist in humans. Since there are data supporting

the existence of the 4.4 kb *hcasz5* (GenBank Accession No. AK092289), we sought to further evaluate the existence of the longer *hcasz11* transcript obtained from in silico analysis by another method. We utilized RNA interference to knockdown the expression of *hcasz5* and test the changes in expression level of both *hcasz5* and *hcasz11* by real-time PCR. As shown in Fig. 5A, the siRNA target was located at 5′ terminal of *hcasz5*, while two pairs of primers were designed to distinguish *hcasz5* and *hcasz11* levels. Primer pair 1 was located around the siRNA target sequence, while primer pair 2 was designed to bind to a 3′ terminal location in the putative *hcasz11* but not in the *hcasz5 mRNA*. *Hcasz5* and *hcasz11* have the same sequence from the beginning to PS1 (PS1: pre-stop codon of *hcasz5*) and there is no overlap after that position. After transfection of *hcasz5* siRNA for 48 h, the *hcasz5* mRNA level was reduced nearly 80% as detected by real-time PCR using primer pair 1 (Fig. 5B). We hypothesized that if *hcasz11* does not exist, the mRNA level should not change when primer pair 2 is used for real-time PCR because the target of *hcasz5* siRNA is not on the same mRNA as primer pair 2. However, the results indicate that after the knockdown of *hcasz5* by *hcasz5* siRNA, *hcasz11* levels decreased nearly 60% (Fig. 5B). This result supports the existence of the longer transcript.

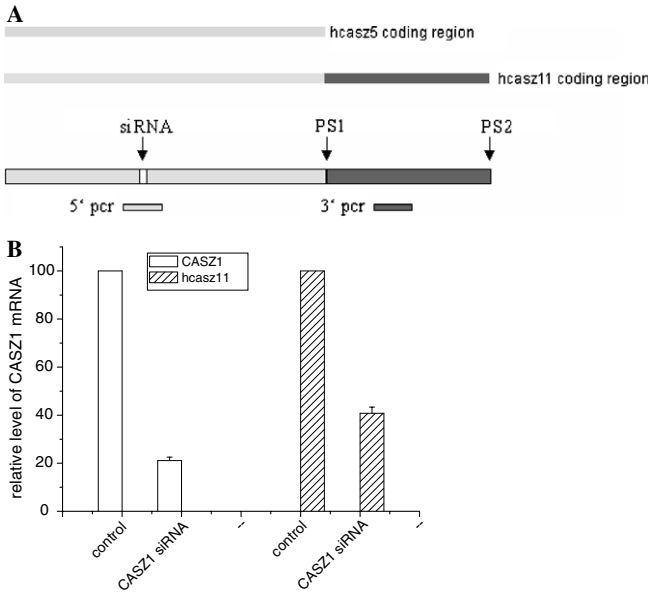

Fig. 5. Confirmation of *hcasz11* transcript. (A) *Hcasz5* and *hcasz11* have the same sequence from the beginning to PS1 (PS1: pre-stop-codon of *hcasz5*) but there is no sequence overlap after that position. A siRNA was designed targeting the 5′ terminal of *hcasz5* that is a region also contained in *hcasz11* and thus would inhibit expression of all *CASZ1* isoforms. Two pairs of primers were designed, one pair located around the siRNA target sequence that is common to both *hcasz5* and *hcasz11*, and the other pair is located at 3′ terminus and is only found in *hcasz11* (PS1, pre-stop codon of *hcasz5*; PS2, pre-stop-codon of *hcasz11*. 5′ pcr: amplify 5′ fragment shown in (A), which belongs to both *hcasz5* and *hcasz11*. 3′ PCR: amplify 3′ fragment shown in (A), which only belongs to *hcasz11*). (B) The change of *CASZ1* mRNA level in AS cells after *hcasz5* siRNA treatment is assessed by real-time PCR. Both *hcasz5* and *hcasz11* are knocked down by the siRNA, and that supports the existence of the longer transcript (primer pair 1: amplify 5′ PCR fragment shown in (A); primer pair 2: amplify 3′ PCR fragment shown in (A)).

The 5′-flanking region of *CASZ1* was analyzed and its function was studied. DNA sequences from +62 to −3398 (−1 represents the first nucleotide upstream of *CASZ1* exon 1) were analyzed by employing TRANSFAC database [13] and MAPPER program [14]. The programs predicted a number of putative response elements, including for N-Myc, c-Myc, CREB, CDP CR1, Pax-6, v-Maf, AP-1, Pax-4, CCAAT box, NF-Y, COMP1, p53, and Oct-1 (Fig. 6). The neural origin cell line SK-N-AS and non-neural origin cell line NIH3T3 were chosen as model systems for *CASZ1* 5′-flanking region analysis. The transcription start point was determined by the 5′ RACE and confirmed to be consistent with the most 5′ end sequence of all *CASZ1* transcripts in GenBank. Based on this result, different length fragments of *CASZ1* 5′-flanking region from +62 to −3298 were cloned upstream of a luciferase reporter gene. Luciferase activity was normalized to β-galactosidase activity, and the pGL2 basic empty vector luciferase activity was set to 1. Transient transfection studies showed our cloned 5′-flanking region was transcriptionally active. In SK-N-AS, five of the promoter constructs have at least a 2.5-fold increase in luciferase activity compared with the empty vector control (Fig. 7). The −18*Cas* pLuc only has 80 bp and has no significant luciferase activity. From Fig. 7 we can see 738*Cas* pLuc still has a 3-fold increase in luciferase activity compared to the empty vector control, which means the promoter of *CASZ1* retains some functional activity within 738 bp 5′ of the start site. The sudden increase in transcriptional activity of −1957*Cas* pLuc compared with −2822*Cas* pLuc corresponds to deletion of putative c-Myc, CDP CR1, Pax6, and CREB binding sites located in the region from −2822 to −1957 bp and implies that some of these transcription factors may function as transcriptional inhibitors. The gradual decrease in transcriptional activity from −3298*Cas* pLuc to −2822*Cas* pLuc, as well as from −1957*Cas* pLuc to −1278*Cas* pLuc and −738*Cas* pLuc, suggests that the following regions: from −3298 to −2822, from −1957 to −1278, from −1278 to −738, and from −738 to −18 may contain positive regulators of transcriptional activity. In the NIH3T3 cell line, a similar pattern of promoter activity was found (Fig. 7), although the relative promoter activity is much lower in NIH-3T3 cells compared to SK-N-AS. This is consistent with Northern blot analysis, which shows little castor expression in NIH3T3. Putative transcriptional factor binding sites are shown in Fig. 6.

To study the protein products of *hcasz5* and *hcasz11*, we obtained an *hcasz5* cDNA clone containing 4406 bp of nucleotide sequence and an *hcasz11* cDNA clone containing 7946 bp of nucleotide sequence. A fusion protein with FLAG epitope attached to the N terminus of *hcasz5* or *hcasz11* protein was made to identify the *hcasz5* or *hcasz11* protein by cloning *hcasz5* and *hcasz11* genes into pCMV Tag 2A vector. An anti-FLAG M2 monoclonal

Fig. 6. Nucleotide sequence of the 5′-flanking region (from −3298 to +62) of the *CASZ1* gene. The first nucleotide of *CASZ1* exon 1 is set as +1 and the nucleotide within exon 1 is shown as lowercase. Putative binding sites for transcription factors are underlined (the sequences underlined with no labels denote c-Myc sites).

antibody was used to identify FLAG-*hcasz5* and FLAG-*hcasz11*. We transfected pCMV-Tag 2A *hcasz5* and pCMV-Tag 2A *hcasz11* into SK-N-AS cell, and Western blot analysis indicated that the apparent molecular weights of *hcasz5* and *hcasz11* in SDS–PAGE analysis approximate the predicted sizes of 125 and 190 kDa for *hcasz5* and *hcasz11*, respectively. Cell fractionation studies indicate that both *hcasz5* and *hcasz11* are predominantly expressed in the nucleus, although a low level of *hcasz5* expression was detected in the cytoplasm (Fig. 8A). This result is confirmed by immunostaining that reveals that both *hcasz5* and *hcasz11* are predominantly localized in the nucleus (Fig. 8B).

Discussion

In this report, we present the cloning and characterization of the mammalian human *castor* gene and identify a novel isoform not previously recognized in the genomic DNA databases. We found that human*CASZ1* shows a broader range of expression than that recognized in *Drosophila* with high expression in heart, lung, skeletal muscle, pancreas, testis, small intestine, and stomach (Fig. 1A). Although *CASZ1* expression was not detected in adult human brain, we found variable *castor* gene mRNA expression in different regions of postnatal day 6 mouse brain (Fig. 1B). A previous report showed that there is

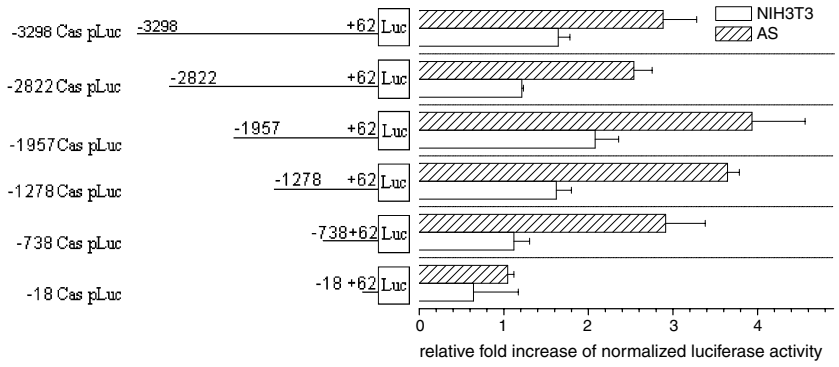

Fig. 7. *CASZ1* promoter analysis. Deletion analysis of the *CASZ1* promoter in SK-N-AS and NIH3T3 cells. Luciferase activities are normalized to cotransfected β-galactosidase and then set relative to the pGL2 basic vector (pGL2 basic vector set as 1).

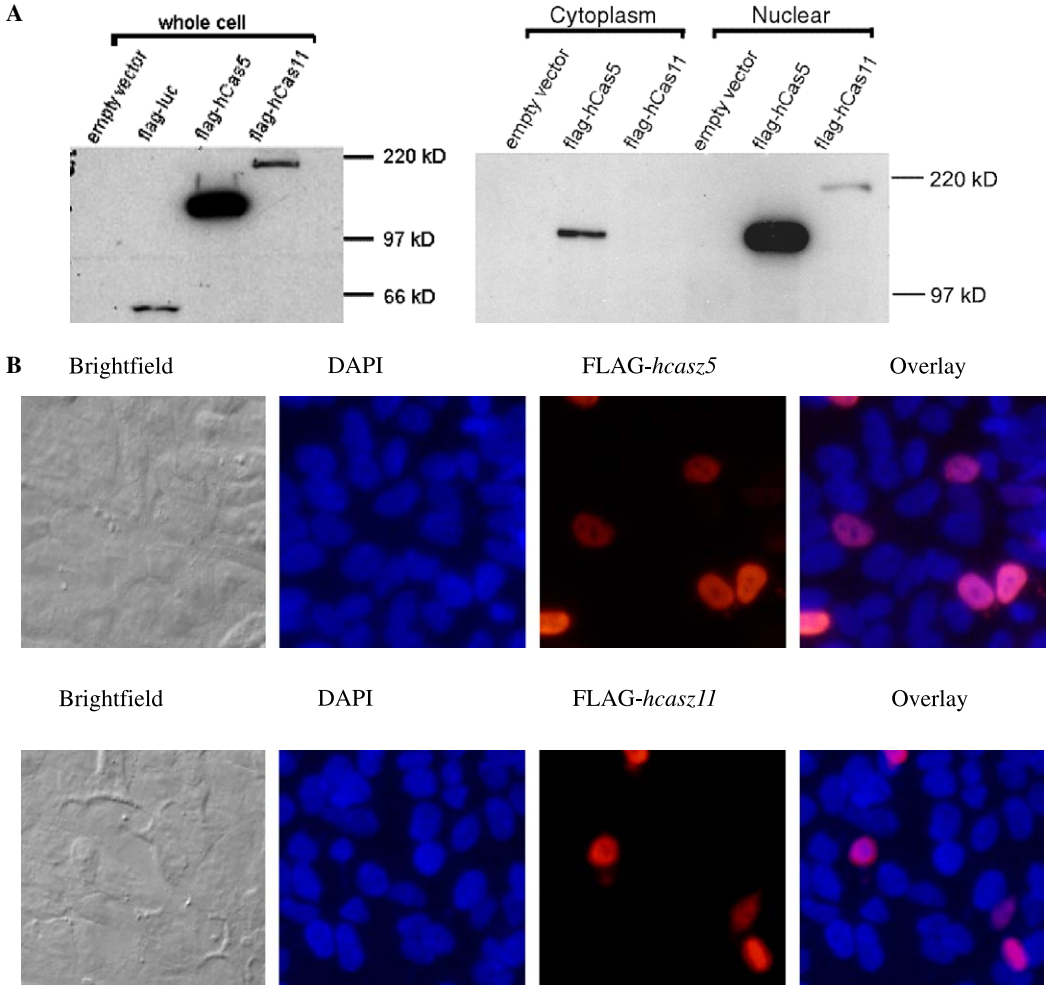

Fig. 8. SK-N-AS cells over-expressing FLAG-*hcasz5* and FLAG-*hcasz11*. *Hcasz5* and *hcasz11* were predominantly expressed in nucleus. (A) Tested by Western blot, the proteins were extracted from the whole cell, cytoplasm, and nucleus. Empty vector represents pCMV-Tag 2A and FLAG-luc represents positive control vector expressing FLAG-tagged luciferase. (B) Tested by Immunocytochemistry, the nuclear localization of FLAG-*hCas* was characterized by the colocalization of rhodamine-labeled goat anti-mouse IgG and the DAPI-stained nucleus.

expression of the mouse *castor* related EST (*Cst*) in parts of the fetal hindbrain, as well as cranial and dorsal root ganglia of the peripheral nervous system during embryogenesis. These studies are consistent with a role for *castor* in regulating neuroblast cell fate in mammals during an early stage of development. Aside from being expressed in the neural system, *Cst* was also detected in the developing heart [20]. This is consistent with our finding that there is a high level of *CASZ1* expression in heart and skeletal muscle as well as in rhabdomyosarcoma, an embryonic tumor

of muscle origin. *CASZ1* is expressed in a variety of different normal tissues such as heart, lung, skeletal muscle, pancreas, testis, small intestine, and stomach, and is expressed in a number of tumor cell lines from neural (retinoblastoma, neuroblastoma, and glioblastoma) as well as non-neural tissues (rhabdomyosarcoma, an embryonal tumor of muscle cells). This indicates that aside from having a role in neural fate determination, *CASZ1* may also be important in developmental programs controlling muscle differentiation and/or the differentiation programs of other tissues.

If *castor* gene expression were involved in tissue development then one would predict that it should be regulated during differentiation. Here, we adopted two cell lines including KCNR and C2C12. KCNR is neural origin and C2C12 is muscle origin, both of them are well-studied models of differentiation with easily detected changes in cell morphology marking the differentiated state, such as neurite extension and myotube formation [18,19]. The results showed that *castor* gene expression was significantly increased in both KCNR and C2C12 when they were induced to differentiation (Fig. 2). This indicates that the *castor* gene expression is expressed in a developmentally regulated manner in mammalian differentiation as it is in *Drosophila* and implies that the *castor* gene may play a central role in controlling cell fates not only within neuroblast cell lineages [1–7], but it may also regulate cell fate in myoblast cell lineages as well.

The *CASZ1* Northern blot analysis indicated the presence of two mRNA species of 4.4 and 8.0 kb. By searching the GenBank, we found the 4.4 kb of *hcasz5* transcript (GenBank Accession No. AK092289). Our in silico analysis led to the discovery of a novel, longer, 8 kb *hcasz11* by utilizing AK092289 as template and the resulting cDNA contained start-, stop-codon, and poly(A) signal (Figs. 4A and B). The relative level of expression of hCas11 is greater than that of hCas5 in most of the tissues and cell lines examined by Northern blot analysis which may indicate that this isoform plays an important functional role.

Protein analysis revealed that *hcasz5* has 5 class 1 C2H2 zinc finger (ZnF) motifs, with highly conserved *para*-ZnF sequence that is unique to *Drosophila castor* (Fig. 1). Moreover, *hcasz11* has 6 more of this same class of ZnF/*para*-ZnF motifs than *hcasz5*. The additional ZnF in *hcasz11* suggests that there may be some functional differences such as different DNA sequence binding or different protein–protein interactions from *hcasz5*. A fusion protein with a FLAG epitope attached to the N terminus of *hcasz5* or *hcasz11* was utilized to evaluate the subcellular location of these proteins. Western blot analysis and immunostaining of transfected cells reveal that *hcasz5* and *hcasz11* predominantly localize to the nucleus as predicted for transcription factors (Fig. 8). The more ZnF and low complexity composites existing in *hcasz11* imply the functional difference of the two proteins, perhaps the *hcasz5* can be a dominant negative product to regulate some of hcasz11 function. Protein sequence homology analysis from different species indicates that the predicted *castor* proteins usually have about 1200 amino acid residues, with approximately 90% identity to *hcasz5*. One exception was *Tetraodon nigroviridis castor*, in which the predicted protein has 1765 amino acids and the protein identity is 63% when compared with *hcasz11*. Since the longer *castor* protein exists in *Tetraodon nigroviridis* and we have identified a longer *castor* protein in *Homo sapiens*, it is also possible that the longer splicing transcript of *castor* encoding a larger protein also exists in other species, as well as in *Drosophila*. From the sequence homology analysis we can see that the *castor* is evolutionarily conserved in divergent species and such evolutionary conservation implies a broad functional importance.

Transient transfection studies show our cloned 5′-flanking region can function to promote the transcription of the luciferase reporter gene (Fig. 7). The sudden increase in transcriptional activity in −1957*Cas* pLuc compared with −2822*Cas* pLuc implies the presence of transcriptional inhibitors in this region, which contains putative c-Myc, CDP CR1, Pax6, and CREB binding sites. The gradual decrease in transcriptional activity from −1975 to −18 suggests the loss of transcriptional activators. Our finding that the proximal 5′-flanking regions of CASZ1 have a similar pattern of expression in both cells of neural origin (AS cell) and in fibroblast cells (NIH3T3) lends strong support for this 5′-flanking region's function as a promoter of gene transcription. Future studies are aimed at elucidating the factors that regulate *CASZ1* transcription.

In conclusion, this report details the molecular cloning and genomic analysis of the *CASZ1* gene. Although *CASZ1* had been identified in silico as a human homolog of the *Drosophila* castor gene, there were no reports on the structure and regulation of this gene. The CASZ1 gene has two predominant transcripts, *hcasz5* that was predicted from in silico analyses and the novel isoform we identified in this study, *hcasz11*. Both *hcasz5* and *hcasz11* are selectively expressed in distinct normal tissues/organs and tumor cell lines of different tissue origins, although the relative expression of the hcas11 isoform is greater than that of the hcas5 isoform in most cells. *Drosophila castor*, the homolog of *CASZ1* in *Drosophila*, controling cell fates within neuroblast cell lineages suggests *castor* expression may be important for cell proliferation, development, and differentiation as well as carcinogenesis. Further experiments are ongoing to determine the function of *CASZ1* in regulating gene expression and differentiation.

Acknowledgments

We thank all the members of the Cellular and Molecular Biology Section for their support. This research was supported by the Intramural Research Program of the NIH, National Cancer Institute, Center for Cancer Research.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.bbrc.2006.03.207.

References

[1] X. Cui, C.Q. Doe, Ming is expressed in neuroblast sublineages and regulates gene expression in the *Drosophila* central nervous system, Development 116 (1992) 943–952.
[2] D.M. Mellerick, J.A. Kassis, S.D. Zhang, W.F. Odenwald, Castor encodes a novel zinc finger protein required for the development of a subset of CNS neurons in *Drosophila*, Neuron 9 (1992) 789–803.
[3] R. Kambadur, K. Koizumi, C. Stivers, J. Nagle, S.J. Poole, W.F. Odenwald, Regulation of POU genes by castor and hunchback establishes layered compartments in the *Drosophila* CNS, Genes Dev. 12 (1998) 246–260.
[4] R. Hitier, M. Chaminade, T. Preat, The *Drosophila* castor gene is involved in postembryonic brain development, Mech. Dev. 103 (2001) 3–11.
[5] T. Brody, W.F. Odenwald, Cellular diversity in the developing nervous system: a temporal view from *Drosophila*, Development 129 (2002) 3763–3770.
[6] W.F. Odenwald, Changing fates on the road to neuronal diversity, Dev. Cell 8 (2005) 133–134.
[7] R. Grosskortenhaus, B.J. Pearson, A. Marusich, C.Q. Doe, Regulation of temporal identity transitions in *Drosophila* neuroblasts, Dev. Cell 8 (2005) 193–202.
[8] R. O'Rahilly, F. Muller, Minireview: summary of the initial development of the human nervous system, Teratology 60 (1999) 39–41.
[9] C.J. Thiele, C.P. Reynolds, M.A. Israel, Decreased expression of N-myc precedes retinoic acid-induced morphological differentiation of human neuroblastoma, Nature 313 (1985) 404–406.
[10] X. Yang, M.S. Merchant, M.E. Romero, M. Tsokos, L.H. Wexler, U. Kontny, C.L. Mackall, C.J. Thiele, Induction of caspase 8 by interferon gamma renders some neuroblastoma (NB) cells sensitive to tumor necrosis factor-related apoptosis-inducing ligand (TRAIL) but reveals that a lack of membrane TR1/TR2 also contributes to TRAIL resistance in NB, Cancer Res. 63 (2003) 1122–1129.
[11] T. Matsuo, J.K. Stauffer, R.L. Walker, P. Meltzer, C.J. Thiele, Structure and promoter analysis of the human unc-33-like phosphoprotein gene: E-box required for maximal expression in neuroblastoma and myoblasts, J. Biol. Chem. 275 (2000) 25052.
[12] S.F. Altschul, T.L. Madden, A.A. Schaffer, J. Zhang, Z. Zhang, W. Miller, D.J. Lipman, Gapped BLAST and PSI-BLAST: a new generation of protein database search programs, Nucleic Acids Res. 25 (1997) 3389–3402.
[13] A.E. Kel, E. Gossling, I. Reuter, E. Cheremushkin, O.V. Kel-Margoulis, E. Wingender, MATCH: a tool for searching transcription factor binding sites in DNA sequences, Nucleic Acids Res. 31 (2003) 3576–3579.
[14] V.D. Marinescu, I.S. Kohane, A. Riva, MAPPER: a search engine for the computational identification of putative transcription factor binding sites in multiple genomes, BMC Bioinformatics 6 (2005) 79.
[15] B. Morgenstern, DIALIGN 2: improvement of the segment-to-segment approach to multiple sequence alignment, Bioinformatics 15 (1999) 211–218.
[16] L. Falquet, M. Pagni, P. Bucher, N. Hulo, C.J. Sigrist, K. Hofmann, A. Bairoch, The PROSITE database, its status in 2002, Nucleic Acids Res. 30 (2002) 235–238.
[17] I. Letunic, R.R. Copley, S. Schmidt, F.D. Ciccarelli, T. Doerks, J. Schultz, C.P. Ponting, P. Bork, SMART 4.0: towards genomic data integration, Nucleic Acids Res. 32 (2004) D142–D144.
[18] S. Burattini, P. Ferri, M. Battistelli, R. Curci, F. Luchetti, E. Falcieri, C2C12 murine myoblasts as a model of skeletal muscle development: morpho-functional characterization, Eur. J. Histochem. 48 (2004) 223–233.
[19] C. Gaetano, K. Matsumoto, C.J. Thiele, In vitro activation of distinct molecular and cellular phenotypes after induction of differentiation in a human neuroblastoma cell line, Cancer Res. 52 (1992) 4402–4407.
[20] C.M. Vacalla, T. Theil, Cst, a novel mouse gene related to *Drosophila* Castor, exhibits dynamic expression patterns during neurogenesis and heart development, Mech. Dev. 118 (2002) 265–268.