Biochemical and Biophysical Research Communications 435 (2013) 160–164

Contents lists available at SciVerse ScienceDirect

Biochemical and Biophysical Research Communications

journal homepage: www.elsevier.com/locate/ybbrc

Identification, expression and subcellular localization of *ESRG*

Guifei Li [a], Caiping Ren [a,*], Jia Shi [a], Wei Huang [a], Hui Liu [a], Xiangling Feng [a], Weidong Liu [a], Bin Zhu [a], Chang Zhang [a], Lei Wang [a], Kaitai Yao [a,c], Xingjun Jiang [b,*]

[a] Cancer Research Institute, Xiang-Ya School of Medicine, Central South University, Key Laboratory for Carcinogenesis of Chinese Ministry of Health, Key Laboratory for Carcinogenesis & Cancer Invasion of Chinese Ministry of Education, Changsha, Hunan 410078, PR China
[b] Department of Neurosurgery, Xiangya Hospital, Central South University, 87 Xiangya Road, Changsha, Hunan 410008, PR China
[c] Cancer Research Institute, Southern Medical University, 1838 Guangzhou N. Rd., Guangzhou, Guangdong 510515, PR China

ARTICLE INFO

Article history:
Received 16 April 2013
Available online 27 April 2013

Keywords:
ESRG
Human embryonic stem cell
cDNA cloning
Cellular localization
mRNA expression

ABSTRACT

ESRG (embryonic stem cell related gene, also known as *HESRG*), is a novel human gene first cloned and identified by our group with microarray analysis. Interestingly, it is expressed specifically in undifferentiated human embryonic stem cells (hESCs), while its expression pattern and its role in hESCs remain unclear. Here, full-length 3151nt *ESRG* cDNA was further identified by RNA ligase mediated rapid amplification of cDNA ends (RLM-RACE) technique. Meanwhile, an alternatively splicing *ESRG* transcript (*ESRG-B*) of 2837nt in length was also found. Surprisingly, bioinformatics analyses showed that the open reading frames (ORFs) of *ESRG* and *ESRG-B* were identical. Both of them consist of 669nt and encode a 222aa protein with a predicted molecular size of 24 kDa. The ESRG protein was located in the nuclei of hESCs as demonstrated by immunocytochemical staining and Western blotting using ESRG specific antibody generated by us. In contrast, ESRG located in the cytoplasm of COS7 cells when it was forced to be expressed in these cells by gene transfection strategy, suggesting there may be some special proteins present only in hESCs which can help ESRG protein transport into the nuclei of hESCs. By spatial expression analysis, we further discovered that *ESRG* only expressed in the ovary tissue and hESCs instead of other tissues or cell lines. Our current data provide us with an important basis for conducting further studies on the functions and regulatory mechanisms underlying the role of *ESRG* in hESCs.

1. Introduction

Embryonic stem cells (ESCs) are pluripotent stem cells derived from the inner cell mass of a pre-implantation blastocyst, an early-stage embryo. ESCs can be induced to differentiate into different cell lineages *in vitro* [1]. Thus, it provides us an experimentally tractable *in vitro* system to understand the molecular programs that control the earliest steps of lineage specification and maintenance of pluripotency. Oct4, Sox2, and Nanog are often described as the core set of transcription factors necessary for maintaining the pluripotency of ESCs [2]. However, not all the genes related to the pluripotency of ESCs have been discovered [3,4].

ESRG, a novel human embryonic stem cell related gene, is located at chromosome 3p14.3 and composed of four exons and three introns with a full-length mRNA of 3151 nucleotides (nt). It only expressed in undifferentiated hESCs but not in their differentiated ones [5]. Presumably, it may play an important role in the maintenance of self-renewal or pluripotency of hESCs, possibly through involvement in the core regulatory network consisting of *Oct4*, *Sox2*, *Nanog*, *etc*. Bioinformatics analysis shows that *ESRG* gene may contain an open reading frame (ORF) comprising 669nt with a CTG initiation codon at 2–4nt and a stop codon at 668–670nt, and encode a 24 kDa protein containing 222 amino acid (aa) residues. It is predicted to be located in the nuclei. In this report, we used expression strategy to identify the ORF and subcellular localization of ESRG. Meanwhile, we detected *ESRG* expression in various tissues to profile its spatial expression pattern. As far as we know, this is the first report to show the ORF and subcellular localization of ESRG, and our current data provides us with an important basis for conducting further studies on the special role of *ESRG* in hESCs and the underlying molecular mechanisms.

1.1. Cell culture

Human ES cells (H9 and H1 cell lines) were from WiCell Research Institute (Madison, WI, USA) and cell culture was performed as described previously [6,7]. Briefly, hESCs were cultured on mitomycin-C treated mouse embryonic fibroblast (MEF) feeder layer in DMEM/F12 medium (Invitrogen, USA) supplemented with 20% knockout serum replacement (KSR) (a serum-free formulation)

* Corresponding authors.
E-mail addresses: rencaiping@csu.edu.cn (C. Ren), jiangxingjun@sina.com (X. Jiang).

0006-291X/$ - see front matter © 2013 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.bbrc.2013.04.062

(Invitrogen), 1 mM glutamine (Invitrogen), 0.1 mM β-mercaptoethanol (Sigma, USA), 1% nonessential amino acids (Invitrogen), and 8 ng/ml basic fibroblast growth factor (bFGF, PeproTech, USA). To obtain a feeder-free culture, the cells were cultured on Matrigel (BD Bioscience, USA) and grown in media conditioned for at least 24 h by MEFs. C57BL/6 mouse ES cell line (ATCC, USA) was maintained in DMEM (Invitrogen) supplemented with 15% FBS (Gibco), 0.1 mM β-mercaptoethanol, 1 mM glutamine, 1% nonessential amino acids and 1,000 units/ml of leukemia inhibitory factor (LIF) (Millipore, USA). All regular media were supplemented with 10% FBS for other cell lines. SK-OV-3 (human ovarian carcinoma cell line) was cultured in DMEM/F12 medium (Invitrogen). HeLa (human cervical carcinoma epithelial cell line) was grown in minimum essential medium (MEM) (Hyclone, USA). HEK-293 (human embryonic kidney cell line), COS7 (African green monkey kidney cell line) and NIH/3T3 (mouse embryonic fibroblast cell line) were grown in DMEM. JAR (human choriocarcinoma cell line) was grown in RPMI 1640 (Invitrogen). These cell lines were kept in our lab.

1.2. Tissue samples

Seven kinds of human fetal tissues (including skin, lung, brain, thyroid, ovary, thymus, liver) of aborted fetus (36 weeks old) and adult testis tissue were obtained with informed content and the study was performed with approval of the ethics committee of each institution involved in this project.

1.3. 5′- and 3′-rapid amplification of cDNA ends (RACE)

Prior to this report, the 5′-end of ESRG was obtained using SMART RACE cDNA Amplification Kit (Clontech). However, the main limitation of this classic RACE technique is that there is no selection for amplification of fragments corresponding to the actual 5′-ends of mRNA: all cDNAs are acceptable templates in polymerase chain reaction (PCR). RNA ligase mediated rapid amplification of cDNA ends (RLM-RACE) is designed to amplify cDNA only from full-length, capped mRNA, usually producing a single band after PCR, so its result is more reliable [8].

Thus, we amplified the 5′-end of ESRG using a commercially available RLM-RACE kit (GeneRacer™ Kit, Invitrogen). Briefly, 2–5 μg DNase I-treated RNA was dephosphorylated with calf intestine phosphatase, digested by tobacco acid pyrophosphate (TAP) to remove the 5′cap structure, and ligated to 5′ RACE adapter at 5′-end using T4 RNA ligase. The ligated RNA was transcribed into cDNA with random primer and used as a template for the PCR reaction with GeneRacer 5′primer and 5′-GSP (Table 1). Each PCR was run in 50 μl of reaction volume containing 5 μl of 10 × PrimeSTAR™ buffer (Mg^{2+} plus), 4 μl of dNTPs mixture (2.5 mM each), 1 μl of each primers (10 mM), 2 μl of 5′-RACE-Ready cDNA, 0.5 μl PrimeSTAR™ HS DNA polymerase (2.5 U/μl) and 36.5 μl PCR-grade water. The touchdown PCR conditions were 94 °C for 3 min, 5 cycles (94 °C for 30 s, 72 °C for 1 min), 5 cycles (94 °C for 30 s, 70 °C for 1 min), 22 cycles (94 °C for 30 s, 68 °C for 30 s, 72 °C for 1 min), and then a final extension at 72 °C for 10 min.

The 3′-end of the cDNA was amplified by SMART RACE cDNA Amplification Kit (Clontech). Briefly, 2–5 μg DNase I-treated RNA was transcribed into cDNA with 3′ RACE adapter and used as a template for the PCR reaction. The touchdown PCR was performed with UPM (Universal Primer Mix) and 3′-GSP (Table 1) under the same condition as 5′ RACE PCR.

Both PCR products for 5′-end and 3′-end were electrophoresed in a 1.5% agarose gel and purified from the gel using Qiaquick Gel Extraction Kit (Qiagen, Germany). The purified PCR products were subcloned into the pMD18-T vector (TaKaRa, Japan). Plasmid DNA was extracted with Plasmid Mini Purification Kit (Qiagen) and sequenced by Invitrogen Biotech Co. Ltd. (Shanghai, China). Two primers complementary to the T7 and SP6 promoters of the vector were used for sequencing. Ten independent clones of each PCR product were analyzed to avoid errors in sequence analysis.

The complete sequence of the *ESRG* cDNA was deduced from the overlapping sequences of both 5′-end and 3′-end amplification products. All oligonucleotides used in this study were synthesized by Invitrogen Biotech Co. Ltd..

Table 1
Primers used in this study.

Primer name	Primer sequence (5′ to 3′)
RACE-PCR primers	
5′-GSP	TAGTGAGGGAGGTTGGAGAACAGAGTA
GeneRacer 5′primer	CGACTGGAGCACGAGGACACTGA
3′-GSP	AAGCAGTGGTATCAACGCAGAGT
UPM	ctaatacgactcactatagggcAAGCAGTGGTATCAACGCAGAGT
ORF cloning primers	
ORF sense	CTGACTCTCTTTTCGGACTCAG
ORF antisense	TGAAAATAAGCGATTGGGGGGTT
Gene-specific primers for RT-PCR	
GAPDH forward	CTTTGGTATCGTGGAAGGACTC
GAPDH reverse	CTCTTCCTCTTGTGCTCTTGCT
ESRG forward	ATGAAAGGGAAGACATACAA
ESRG reverse	TGAACATAGCAAGGGAAA
ESRG-B forward	TGAAAGGGAAGACATACAAAAAC
ESRG-B reverse	GAACATCTCCAGAACAACCTCACAG

1.4. Sequence analysis

The similarity analyses of the determined nucleotide sequences and deduced amino acid sequences were performed by BLAST programs <http://blast.ncbi.nlm.nih.gov/Blast.cgi>. The prediction of transcription start sites was conducted using TSS finder <http://www.fruitfly.org/seq_tools/promoter.html> [9]. The possible ORF was analyzed with ORF Finder tool <http://www.ncbi.nlm.nih.gov/gorf/gorf.html>. The protein sequence of the cloned gene, its molecular weight, pI and topology predictions were conducted with the Expert Protein Analysis System <http://expasy.org/tools/> [10]. The simple modular architecture research tool (SMART) version 4.0 <http://smart.embl-hidelberg.de/> was used to predict the protein domains [11]. The subcellular localization of ESRG protein was predicted using ngLOC software <http://ngloc.unmc.edu/index.html> [12].

1.5. Reverse transcription-PCR (RT-PCR)

All PCR primers used in this study were summarized in Table 1. Total RNA was extracted with Trizol reagent (Invitrogen) and digested by the RNase-free DNase I (TaKaRa) to remove trace amounts of genomic DNA contamination. The concentration and quality of total RNA were determined by UV-absorbance at 260 nm, the A260/A280 ratio and agarose gel electrophoresis. The first-strand cDNA was synthesized from total RNA using reverse transcription system (Promega, USA) and subjected to PCR (95 °C for 5 min; 30 cycles through 95 °C for 30 s, 56 °C for 30 s, 72 °C for 30 s, and then extension at 72 °C for 10 min) using DreamTaq Green PCR Master Mix (Thermo, USA) with primers listed in Table 1. *GAPDH* was used as an internal loading control. Controls without template were carried out by replacing cDNA with water. PCR products were analyzed in 1.5% agarose gels and visualized by ethidium bromide staining.

1.6. Construction of ESRG expression vectors

DNA fragment of 2–722 bp containing the ORF of *ESRG* was amplified from H9 cells' cDNA using specific primers (Table 1). Amplification conditions were as follows: initiation at 95 °C for 5 min; 35 cycles of 94 °C for 30 s, 62 °C for 30 s, and 72 °C for 50 s; and a final extension at 72 °C for 10 min. The PCR product was gel purified, cloned into pEASY-T1 vector and sequenced to ensure that no mutations were introduced during PCR amplification. Then the plasmid was digested with *Bam*H I and *Xba* I and cloned into pEF1/His C expression vector (Invitrogen) to obtain the ESRG expression vector, pEF1/His-ESRG. Empty pEF1/His C vector was used as a control and was designated as pEF1/His-Empty. The recombinant plasmid was sequenced to confirm that the inserted sequence was in the proper reading frame.

1.7. Cell transfection

COS7 cell line not expressing the wild type ESRG was used for transfection experiments. A total of 1×10^6 cells were seeded in wells of 6-well plates 24 h prior to transfection. Then the cells were washed with phosphate-buffered saline (PBS; pH 7.2) and the medium was replaced with Opti-MEM (Invitrogen). 2 μg of pEF1/His-ESRG or pEF1/His-Empty was transfected into cells over 6 h using the Fugene HD reagent (Roche, Swiss) according to the manufacturer's instructions. On the next day, the cells were harvested followed by extraction of total protein for Western blotting.

1.8. Production of polyclonal antibodies

We selected an immunogenic epitope on ESRG protein for highest probability of antigenicity using the JW method. Then, we synthesized the specific peptide, coupled it with keyhole limpet hemocyanin (KLH) at C-terminal cysteine residue, and used it to immunize New Zealand rabbits four times each. The rabbit immune serum was collected and polyclonal antibody was purified by affinity chromatography. The antibody titer was measured through enzyme-linked immunosorbent assay (ELISA). The reactivity and specificity of this polyclonal antibody was tested on a variety of cell types. We subsequently used the antibody to detect ERSG expression in hESCs and ESRG-transfected COS7 cells by immunocytochemistry and Western blotting.

1.9. Western blotting

HESCs were lysed in M-PER® Mammalian Protein Extraction Reagent (Thermo). Nuclear and cytoplasmic proteins were extracted using NE-PER® Nuclear and Cytoplasmic Extraction Reagents (Thermo). 1 × Halt™ Protease Inhibitor Cocktail (Thermo) were added to the extracts. Protein concentration was determined using BCA assay and equal amounts of protein from each lysate were run on gradient 4–12% Tris–HCl gel (BioRad, USA). Proteins were transferred to a polyvinylidene difluoride (PVDF) membrane (Millipore), and the membrane was treated with anti-ESRG antibody (diluted 1:300) and anti-β-actin antibody (diluted 1:5000; Sigma) at 4 °C overnight. Secondary antibodies conjugated with horseradish peroxidase (Millipore) were then applied for 1 h at room temperature. The blot was visualized using ECL reagent (Pierce, USA) and recorded with X-ray films.

1.10. Indirect fluorescent immunocytochemistry

The cells were washed with cold PBS and fixed with 4% paraformaldehyde/PBS for 10 min, washed twice with PBS, permeabilized for 10 min with 0.1% Triton X-100, washed three times with PBS, and incubated in 1% BSA for 1 h (to block non-specific protein binding sites) and in rabbit anti-ESRG antibody (diluted 1:150 in 1% BSA) overnight at 4 °C. The cells were then washed three times with PBS and subsequently incubated with the secondary antibody (goat anti-rabbit IgG, fluorescein isothiocyanate (FITC)-conjugated; diluted 1:500) for 1 h at room temperature in the dark. To show the nuclei, cells were further incubated with 10 μM Hoechst 33342 for 10 min. Background staining was examined using the secondary antibody alone. Cells were washed finally with PBS and recorded with inverted fluorescence microscope (Nikon, Japan). Green fluorescence displayed the distribution of ESRG protein, and the cell nuclei were indicated by the blue fluorescence of Hoechst 33342.

2. Results

2.1. Identification and sequence analysis of ESRG and ESRG-B

By overlapping the 3′- and 5′-RACE fragments, a 3151 bp (named *ESRG*) and a 2837 bp (named *ESRG-B*) transcripts were obtained. The only difference between the two was that *ESRG-B* had a truncated 3′UTR compared with *ESRG* (Fig. 1A). In the 3′UTR, three and two mRNA instability motifs (ATTTA) were found in *ESRG* and *ESRG-B*, respectively [13]. The polyadenylation signal (AATAAA) lied at 127 bp upstream of polyA tail. Both transcripts were predicted to contain the same candidate ORF analyzed by ORF Finder on NCBI website. The ORF comprising 669 bp with an initiation codon at 2–4nt and a stop codon at 668–670nt encoded a protein of 222aa with theoretical pI of 9.37 and molecular weight of 24 kDa (Fig. 1B). Furthermore, through comparing our obtained *ESRG* mRNA sequence with the data in GenBank by BLAST, we found a single nucleotide polymorphism (SNP) site at position 374 of *ESRG* ORF (T/C), which led to a shift of S125P (serine to a phenylalanine at 125). But its biologic significance remained unknown. ESRG protein was predicted to be located in nuclei by ngLOC software. SMART analysis showed that it contained a leucine zipper domain (180LAASPAFLGQGQVPLNPFSFTL201). Meanwhile, five protein kinase C phosphorylation sites, one casein kinase II phosphorylation site and two N-myristoylation sites were observed in this protein (Fig. 1B).

2.2. Expression of ESRG and ESRG-B in various cells and tissues

To investigate the mRNA expression pattern of *ESRG* and *ESRG-B*, the semiquantitative RT-PCR method was employed using gene-specific primers (Table 1) with *GAPDH* as internal control. Both *ESRG* and *ESRG-B* were found to be constitutively expressed in hES cell lines H1 and H9, but were undetectable in other cell lines including C57BL/6 mESC, HUVEC, NIH/3T3, HEK-293, COS7, HeLa, JAR, and SK-OV-3 (Fig. 2A). In addition, our results showed that *ESRG* and *ESRG-B* were only weakly expressed in the fetal ovary, and not in other normal tissues including fetal skin, lung, brain, thyroid, thymus, liver and adult testis tissues (Fig. 2B). Both in hESCs and fetal ovary, *ESRG-B* was more abundant than *ESRG* (Fig. 2).

2.3. Subcellular localization of ESRG protein

Western blotting results (Fig. 3A) showed that the affinity-purified rabbit antiserum against ESRG specifically detected a 24 kDa band in hESC extracts and COS7 cell line transfected with pEF1/His-ESRG, which also confirmed the specificity of anti-ESRG. To gain some information on the subcellular localization of ESRG protein, the deduced amino acid sequence was analyzed by ngLOC software. According to the prediction result, ESRG may be located in the nuclei. By isolating cellular different components, we further confirmed that ESRG was specifically localized in the nuclei of

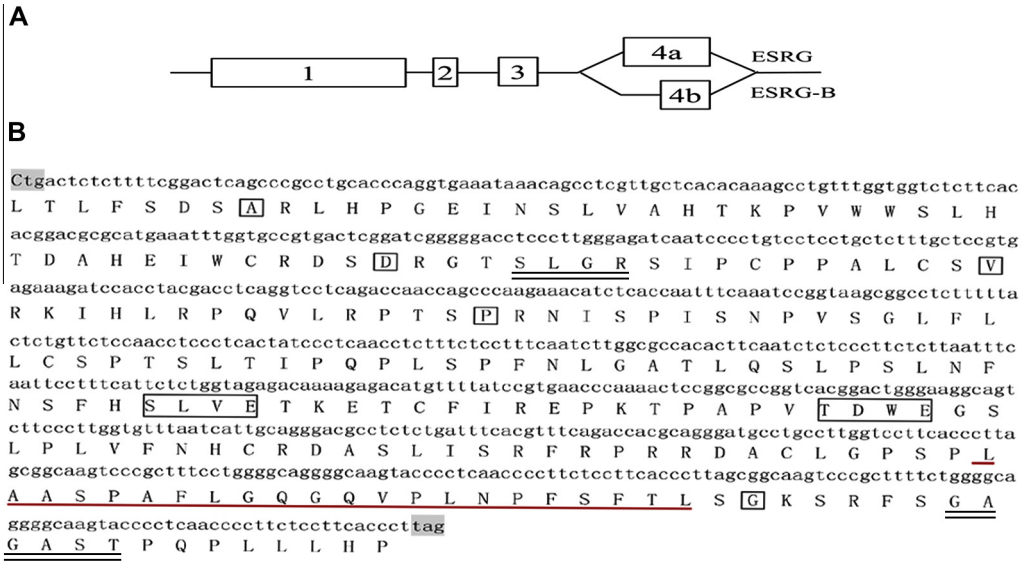

Fig. 1. (A) Comparison of exons between *ESRG* and *ESRG-B*. *ESRG* has exon 4a, while *ESRG-B* has a truncated exon 4b. (B) The nucleotide and deduced amino acid sequences of ESRG. The start (CTG) and stop (TAG) codons are shaded in grey. The underlined amino acids indicate the predicted leucine zipper domain. Five protein kinase C phosphorylation sites and two casein kinase II phosphorylation sites are boxed. Two N-myristoylation sites are double underlined.

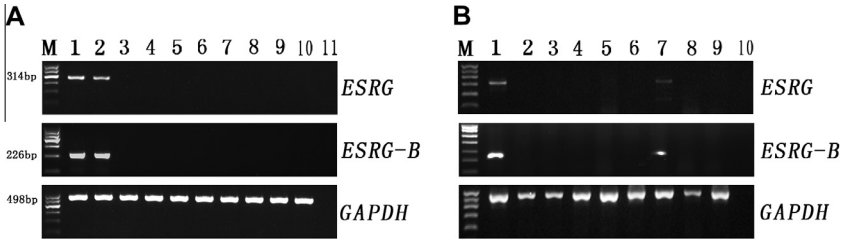

Fig. 2. (A) *ESRG* expression in cell lines. 1, H1 hESCs; 2, H9 hESCs; 3, mESCs; 4, HUVEC; 5, NIH/3T3; 6, HEK-293; 7, COS7; 8, HeLa; 9, JAR; 10, SK-OV-3; 11, no template control (NTC). (B) *ESRG* expression in tissues. 1, H9 hESCs; 2, skin; 3, lung; 4, brain; 5, thyroid; 6, adult testis; 7, ovary; 8, thymus; 9, liver; 10, NTC.

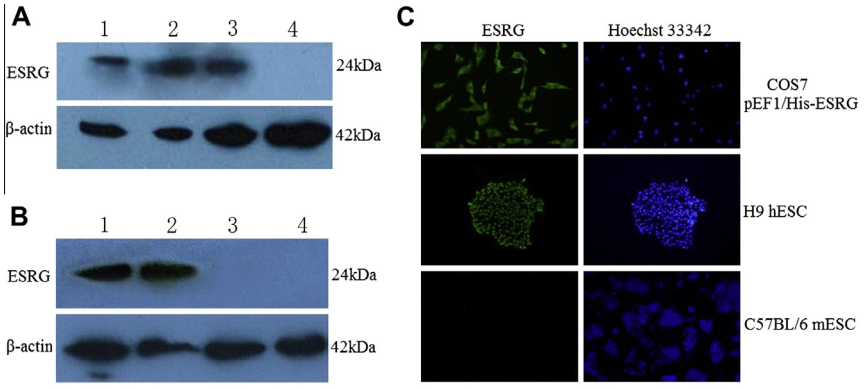

Fig. 3. Subcellular localization of ESRG in hESCs. (A) Western blotting results of ESRG in COS7 cells transfected with pEF1/His-ESRG. 1, H9 hESCs; 2, COS7 cells transfected with pEF1/His-ESRG 24 h post transfection; 3, COS7 cells transfected with pEF1/His-ESRG 48 h post transfection; 4, COS7 cells transfected with pEF1/His-empty vector. (B) Western blotting results of ESRG in nuclear and cytoplasmic extracts from H9 hESCs. 1, H9 total extracts; 2, H9 nuclear extracts; 3, H9 cytoplasmic extracts; 4, mESCs total extracts. (C) Immunostaining of COS7 cells transfected with pEF1/His-ESRG (upper row), H9 hESCs (middle row) and mESCs (lower row) using anti-ESRG (Green). The cells were counterstained for DNA with Hoechst 33342 (Blue).

hESCs (Fig. 3B). Surprisingly, unlike the circumstances in hESCs, when COS7 cell line was forced to express ESRG protein through transfection of pEF1/His-ESRG vector, only cytoplasmic ESRG localization was observed (Fig. 3C) as obviously revealed by immunocytochemistry method. In accordance with above RT-PCR results, we couldn't detect ESRG protein in mESCs by Western blotting or immunocytochemistry (Fig. 3B and C).

3. Discussion

In the present study, for the first time, we have identified and characterized two alternative transcripts of *ESRG* in hESCs. Interestingly, both of them have the same ORF and may encode a same protein. According to the semi-quantitative RT-PCR results, what is intriguing is that it seemed *ESRG-B* was more abundant than *ESRG*

in hESCs. In accordance with the observation, respectively three and two mRNA instability motifs (ATTTA) were found in the 3′UTR of *ESRG* and *ESRG-B* through bioinformatics analysis. The one more mRNA instability motif might be related to stability difference of these two transcripts. These two transcripts were resulted from two 3′ splice sites (*ESRG*, 4a, CATTGCTTATTCCAG) (*ESRG*-B, 4b, TGGTCTTCTCAACAG). The 3′ splice site of *ESRG*-4a appeared to have a poor match with the splicing consensus sequence, i.e. *ESRG*-4a might contain a weaker 3′ splice site than *ESRG-B*-4b [14]. Differences in mRNA instability motif and 3′ splice sites may contribute to the different abundance levels, though further evidence is needed. Also, the biological significance of these two different transcripts in hESCs is worth further investigation.

To understand the function of a novel gene, a good method is to describe its expression pattern in various tissues and cell lines. First, we detected expression of *ESRG* and *ESRG-B* in different cell lines. As we expected, we couldn't detect *ESRG* and *ESRG-B* expression in any cell lines except in hESCs, which is further evidence that *ESRG* is a human embryonic stem cell specific gene. We collected fetal skin, lung, brain, thyroid, ovary, thymus, liver and adult testis tissues to detect *ESRG* expression. Only weak expression of *ESRG* and *ESRG-B* was observed in the ovary, and even in the ovary, *ESRG-B* manifested a higher level than *ESRG*, which indicated that the expression patterns of *ESRG*'s two different mRNA splicing forms were highly similar. Till now, except in hESCs [5], *ESRG* is found to be expressed in only two kinds of cranial germ cell tumors (GCTs) (i.e. germinoma and embryonal carcinoma) [15], which means *ESRG* is one of those undiscovered genes co-expressed among germinoma, embryonal carcinoma (EC) and hESCs. The co-expression of *ESRG* reveals that there might exist some coherent characteristics among these three types of cells. In CNS GCTs, germinoma and EC are thought to be closer to germ cells than other types of GCTs. EC may give rise to embryonal neoplasm with the potential to differentiate into derivatives of all the three germ layers and is considered to be a neoplastic counterpart of hESCs or primordial germ cells (PGCs) [15]. In the ovary, there also exist germ (egg) cells. We speculate that these germ cells may be the *ESRG* positive cell source. From this perspective, another kind of GCT disease related to malignant (cancer) cells formed in the germ (egg) cells of the ovary is worthy of further study.

We subsequently examined the subcellular localization of ESRG in hESCs. Using specific anti-ESRG, as revealed by immunocytochemistry and Western blotting results, we found ESRG protein was localized in the nuclei of hESCs. This was in consistent with bioinformatics prediction results.

We also examined the subcellular localization of exogenously expressed ESRG in COS7 cells by gene transfection strategy. To our surprise, unlike its nuclear localization in hESCs, ESRG protein was localized in the cytoplasm of COS7 cells. In the bioinformatics analysis, we could not find nuclear localization signal (NLS) in ESRG protein, but we found a conserved C terminal leucine zipper domain between amino acids 180 and 201. It is known that many important transcription factors, such as GCN4, c-Fos and c-Jun, and C/EBP etc., have leucine zipper domains [16–18] and this domain is usually involved in homo- or heterodimer formation [19], which then facilitates the interaction of an adjacent region of the protein with specific sequences of the DNA. How ESRG protein, without an NLS site, is transported into the nuclei of hESCs and whether ESRG protein is facilitated by an interaction with another NLS-containing protein via the C-terminal leucine zipper domain remain to be further investigated [20].

In summary, this is the first report for identifying and characterizing *ESRG*'s two alternative mRNA splicing forms, profiling its spatial expression patterns, and analyzing its subcellular localization. The obtained recombinant ESRG protein expression from its cDNA sequence paves the way for subsequent structure–function analysis by mutational studies, and the successful preparation of anti-ESRG antibody will facilitate further studies on the functions and regulatory mechanisms underlying the role of *ESRG* in hESCs.

Acknowledgments

This work was financially supported by National Natural Science Foundation of China (30871246, 81070993), Specialized Research Fund for the Doctoral Program of Higher Education of China (SRFDP) (20120162110059), National Basic Research Program of China (2010CB833605), Program for New Century Excellent Talents in University (NCET-10-0790), Key Program of Central South University (2010QYZD006), Incubation Program for National Natural Science Funds for Distinguished Young Scholar, Open-End Fund for the Valuable and Precision Instruments of Central South University (CSUZC2012010, CSUZC2012011).

References

[1] J.A. Thomson, J. Itskovitz-Eldor, S.S. Shapiro, M.A. Waknitz, J.J. Swiergiel, V.S. Marshall, J.M. Jones, Embryonic stem cell lines derived from human blastocysts, Science 282 (1998) 1145–1147.
[2] L.A. Boyer, T.I. Lee, M.F. Cole, S.E. Johnstone, S.S. Levine, J.P. Zucker, M.G. Guenther, R.M. Kumar, H.L. Murray, R.G. Jenner, D.K. Gifford, D.A. Melton, R. Jaenisch, R.A. Young, Core transcriptional regulatory circuitry in human embryonic stem cells, Cell 122 (2005) 947–956.
[3] M. Richards, S.P. Tan, W.K. Chan, A. Bongso, Reverse serial analysis of gene expression (SAGE) characterization of orphan SAGE tags from human embryonic stem cells identifies the presence of novel transcripts and antisense transcription of key pluripotency genes, Stem Cells 24 (2006) 1162–1173.
[4] J. Yu, J.A. Thomson, Pluripotent stem cell lines, Genes Dev. 22 (2008) 1987–1997.
[5] M. Zhao, C. Ren, H. Yang, X. Feng, X. Jiang, B. Zhu, W. Zhou, L. Wang, Y. Zeng, K. Yao, Transcriptional profiling of human embryonic stem cells and embryoid bodies identifies HESRG, a novel stem cell gene, Biochem. Biophys. Res. Commun. 362 (2007) 916–922.
[6] C. Ren, M. Zhao, X. Yang, D. Li, X. Jiang, L. Wang, W. Shan, H. Yang, L. Zhou, W. Zhou, H. Zhang, Establishment and applications of epstein-barr virus-based episomal vectors in human embryonic stem cells, Stem Cells 24 (2006) 1338–1347.
[7] M. Zhao, H. Yang, X. Jiang, W. Zhou, B. Zhu, Y. Zeng, K. Yao, C. Ren, Lipofectamine RNAiMAX: an efficient siRNA transfection reagent in human embryonic stem cells, Mol. Biotechnol. 40 (2008) 19–26.
[8] B.C. Schaefer, Revolutions in rapid amplification of cDNA ends: new strategies for polymerase chain reaction cloning of full-length cDNA ends, Anal. Biochem. 227 (1995) 255–273.
[9] U. Ohler, H. Niemann, Identification and analysis of eukaryotic promoters: recent computational approaches, Trends Genet. 17 (2001) 56–60.
[10] E. Gasteiger, A. Gattiker, C. Hoogland, I. Ivanyi, R.D. Appel, A. Bairoch, ExPASy: the proteomics server for in-depth protein knowledge and analysis, Nucleic Acids Res. 31 (2003) 3784–3788.
[11] J. Schultz, F. Milpetz, P. Bork, C.P. Ponting, SMART, a simple modular architecture research tool: identification of signaling domains, Proc. Natl. Acad. Sci. USA 95 (1998) 5857–5864.
[12] B.R. King, C. Guda, NgLOC: an n-gram-based Bayesian method for estimating the subcellular proteomes of eukaryotes, Genome Biol. 8 (2007) R68.
[13] G. Shaw, R. Kamen, A conserved AU sequence from the 3' untranslated region of GM-CSF mRNA mediates selective mRNA degradation, Cell 46 (1986) 659–667.
[14] M. Burset, I. Seledtsov, V. Solovyev, Analysis of canonical and non-canonical splice sites in mammalian genomes, Nucleic Acids Res. 28 (2000) 4364–4375.
[15] S. Wanggou, X. Jiang, Q. Li, L. Zhang, D. Liu, G. Li, X. Feng, W. Liu, B. Zhu, W. Huang, J. Shi, X. Yuan, C. Ren, HESRG: a novel biomarker for intracranial germinoma and embryonal carcinoma, J. Neurooncol. 106 (2012) 251–259.
[16] P.R. Mittl, C. Deillon, D. Sargent, N. Liu, S. Klauser, R.M. Thomas, B. Gutte, M.G. Grutter, The retro-GCN4 leucine zipper sequence forms a stable three-dimensional structure, Proc. Natl. Acad. Sci. USA 97 (2000) 2562–2566.
[17] T.K. Kerppola, T. Curran, Fos-Jun heterodimers and Jun homodimers bend DNA in opposite orientations: implications for transcription factor cooperativity, Cell 66 (1991) 317–326.
[18] T. Maekawa, H. Sakura, C. Kanei-Ishii, T. Sudo, T. Yoshimura, J. Fujisawa, M. Yoshida, S. Ishii, Leucine zipper structure of the protein CRE-BP1 binding to the cyclic AMP response element in brain, EMBO J. 8 (1989) 2023–2028.
[19] W.H. Landschulz, P.F. Johnson, S.L. McKnight, The leucine zipper: a hypothetical structure common to a new class of DNA binding proteins, Science 240 (1988) 1759–1764.
[20] G.R. Hicks, N.V. Raikhel, Protein import into the nucleus: an integrated view, Annu. Rev. Cell Dev. Biol. 11 (1995) 155–188.