Mol Biol Rep (2016) 43:65–72
DOI 10.1007/s11033-015-3939-x

ORIGINAL ARTICLE

Identification of new SOX2OT transcript variants highly expressed in human cancer cell lines and down regulated in stem cell differentiation

Marie Saghaeian Jazi[1] · Nader Mansour Samaei[2,3] · Mostafa Ghanei[4] · Mohammad Behgam Shadmehr[5] · Seyed Javad Mowla[6]

Received: 24 March 2015 / Accepted: 18 December 2015 / Published online: 24 December 2015

Abstract Long non-coding RNAs are manifested as a new paradigm of molecular effectors in a wide range of human diseases. Human *SOX2* overlapping transcript (*SOX2OT*) gene can generate six lncRNA transcript variants which are functionally assumed to be correlated with cellular differentiation and carcinogenesis. However, the circumstances determining expressional and functional differences between SOX2OT transcript variants remain to be explored. Here, we studied the expression of all SOX2OT transcript variants specifically in five human cancer cell lines by real-time RT-PCR. Changes of the new SOX2OT transcript variants expression were measured during the NT2 teratocarcinoma cell line neuronal-like differentiation and were compared to pluripotency regulators, *SOX2* and *OCT4A* gene expressions. Surprisingly, we identified two new SOX2OT transcripts, named SOX2OT-7, SOX2OT-8 which lack exon 8. We discovered that beside active proximal and distal SOX2OT promoters, different cancer cell lines express high levels of some SOX2OT transcript variants differentially by alternative splicing. Significantly, both SOX2OT-7 and SOX2OT-8 are highly expressed in human cancer cell lines coinciding with *SOX2*, one of the pluripotency regulators. Our results revealed that SOX2OT-7 is almost the most abundant form of SOX2OT transcript variants in the examined cancer cell lines particularly in NT2 teratocarcinoma cell line where its expression falls upon neuronal-like differentiation similar to *SOX2* and *OCT4A*. We suggest that at least some of SOX2OT transcripts are significantly associated with cancer and stem cell related pathways.

Keywords Long non-coding RNA · *SOX2* overlapping transcript · Carcinogenesis · Neuronal like differentiation

Background

Although the majority of human genome sequence is informative and transcribed into RNA products [1, 2]; it is mostly consists of non protein coding sequences [3, 4]. While the non-coding RNA repertoire of an organism is considerably related to its evolutionary complexity, protein coding gene numbers increase less proportionally [5].

Long non-coding RNAs (lncRNAs) assign approximately 80 % of the eukaryotic transcriptome [6]. They are long (>200 nucleotides) mRNA-like transcripts which harbor post-transcriptional processing and alternative splicing features [7]. To date, more than 200 human lncRNAs are documented in lncRNA database but most of them are not functionally annotated. The great extent of long non-coding RNAs elicits regulatory functions in controlling gene expression events especially at chromatin remodeling, epigenetics, transcriptional and post-transcriptional processes

✉ Nader Mansour Samaei
n_samaei@yahoo.com; samaei@goums.ac.ir

1 Department of Molecular Medicine, Faculty of Advanced Medical Technologies, Golestan University of Medical Sciences, Gorgan, Iran

2 Stem Cell Research Center, Golestan University of Medical Sciences, Gorgan, Iran

3 Department of Human Genetics, Faculty of Advanced Medical Technologies, Golestan University of Medical Sciences, P.O.Box 4934174611, Gorgan, Iran

4 Chemical Injuries Research Center, Baqiyatallah University of Medical Sciences, Tehran, Iran

5 Tracheal Diseases Research Center, National Research Institute of Tuberculosis and Lung Diseases (NRITLD), Shahid Beheshti University of Medical Sciences, Tehran, Iran

6 Department of Molecular Genetics, Faculty of Biological Sciences, Tarbiat Modares University, Tehran, Iran

[7–12]. It has been postulated that lncRNAs can potentially provide a precursor for other small RNAs [13].

SOX2 overlapping transcript (*SOX2OT*) is a conserved lncRNA which encompasses sex determining region Y-box 2 (*SOX2*) in the same strand [14]. There are evidences supporting the association of *SOX2OT* locus (3q26–3q27) with congenital eye disorders [14, 15]. It is proposed that SOX2OT function could be related to *SOX2*, one of the key regulators of stem cells [14, 16]. *SOX2OT* gene including 10 exonic regions presents complex transcriptional features (Fig. 1a); it has at least two transcription start sites, one located upstream producing *SOX2* distal overlapping transcript (SOX2DOT) variants possessing different 5′ nucleotide sequence. Moreover, alternative splicing leads to the production of six different SOX2OT transcript variants; three distal variants (transcript variants 1–3) and three proximal variants (transcript variants 4–6) (Fig. 1b). The expressions of sox2ot and sox2dot transcript variants are dynamically regulated during mouse and other mammalian embryogenesis [17]. The overexpression of SOX2OT is also reported in human cancer tissues of breast, lung and esophagus carcinomas [16, 18–20].

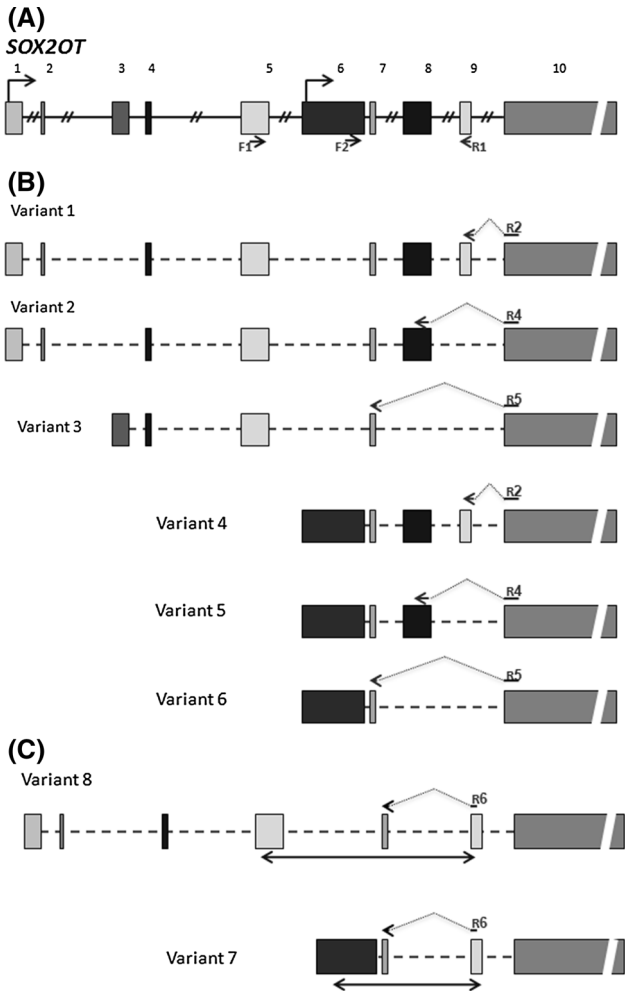


Fig. 1 Schematic representation of *SOX2OT* gene and transcripts. **a** *SOX2OT* gene is regulated by two promoter regions and consists of 10 exons. **b** Schematic reference sequences of six SOX2OT transcript variants are illustrated, notice two groups of distal (1, 2, 3) and proximal (4, 5, 6) variants which are different in 5′ end. Alternative splicing of exons 8 or both 8 and 9 lead to three different 3′ end in proximal variants same as distal transcripts. **c** Schematic representation of the two new discovered SOX2OT-8 and SOX2OT-7 are showed. Notice the new splicing which deletes exon 8. Partially sequenced region of RT-PCR products is demonstrated by *double arrows*. Specific primers binding site used to amplify the transcripts by RT-PCR are indicated by *arrows*

To address the biological importance of SOX2OT alternative splicing process, here for the first time, we investigated all SOX2OT splice variants expressions specifically in five different human cancer cell lines (HeLa, Mcf7, HepG2, U-87 MG and NT2). Surprisingly, another new alternative splicing format, joining exons 7–9 was detected in both distal and proximal transcripts. Along with the previously reported variants, the expression of two novel splice variants, named SOX2OT transcript variant 7 and 8 (GenBanks. KF981435, KJ415055) were also quantified. High expression levels of 4, 7 and 8 variants were detected in all five cancer cell lines; while the rest of SOX2OT splice variants possessed faint expressions, if there was any. *SOX2* gene expression value in cancer cell lines is significantly related to SOX2OT transcript variants 7 and 8 (Pearson's correlation), marking possible biological relevance. Notably, neuronal differentiation of NT2 induce a significant SOX2OT-7 expression decline same as *SOX2* and *OCT4A* which further highlights SOX2OT-7 role in stem cell maintenance regulation.

Methods

Cell culture

Five different human cancer cell lines were selected for this study. HeLa, MCF7, HepG2, U-87 MG were obtained from pasture institute (Tehran, Iran) and NTERA-2 (NT2) cell line was kindly provided by Dr. Peter Andrews (Sheffield University). NT2 and U-87 MG were cultured in DMEM F12 (Invitrogen, Gaithersburg, MD) supplemented with 10 % fetal bovine serum (Invitrogen, Gaithersburg, MD) and 100 IU penicillin-100 µg streptomycin per ml (Invitrogen, Gaithersburg, MD) in a 98 % humidified 5 % CO_2 incubator. For HepG2, HeLa and MCF7 cell lines, RPMI 1640 (Invitrogen, Gaithersburg, MD) used as growth medium in same condition.

Neuronal differentiation

NT2 teratocarcinoma cell line was treated with all-trans retinoic acid (ATRA, Sigma-Aldrich) to differentiate into

the neuron-like cells, according to a modified protocol provided by Atlasi et al. [21]. 3×10^4 NT2 cells were cultured in six-well plates containing 2 ml of growth medium. After 48 h, medium was replaced by differentiation medium containing 10 μM of ATRA. The treatment was continued for 4 weeks whilst differentiation medium was renewed twice a week. Cells were harvested for RNA extraction at time intervals of 0, 3, 7, 14, 21, 32 days.

RNA extraction and reverse transcription

Cells were detached by EDTA/trypsin 0.05 % (Invitrogen, Gaithersburg, MD) then harvested by centrifugation in 4 °C temperature. For total RNA extraction, RNXPlus solution (Cinnaclon, Iran) was used following the manufacturer's instruction. To deplete possible DNA contamination, 1 μg of total RNA were treated with RNase free DNaseI (Thermo Fisher Scientific, Inc) for 30 min in 37 °C followed by DNaseI inactivation using EDTA 5 mM at 65 °C for 10 min. DNA free RNA was reverse transcribed by RevertAid first strand cDNA synthesis kit (Thermo Fisher Scientific, Inc.) using random hexamer primers as described by the supplier. First strand cDNAs were stored at −80 °C until use.

PCR and semi-quantitative real time-PCR

To amplify target transcripts *SOX2* (GenBank: NM_003106.3), SOX2OT transcript variants 1–6 (GenBanks: NR_075091.1, NR_075092.1, NR_075093.1, NR_004053.3, NR_075089.1, NR_075090.1), *OCT4A* (GeneBank: NM_002701.4) and *GAPDH* (GenBank: NM_002046.3); 1 μl of each cDNA was used as template in a polymerase chain reaction of 5 min initial denaturation at 95 °C, 40 cycles of 10 s denaturation at 95 °C, 20 s annealing at 58 °C and 40 s extension at 72 °C, followed by 5 min final extension at 72 °C in Bio-Rad MJ mini thermal cycler. Each 25 μl PCR reaction was composed of 2.5 U taq DNA polymerase (Cinnaclon, Iran), 200 μM dNTPs, 1.5 mM $MgCl_2$ and 5 pmol each primer. All primers (Table 1) were designed specifically by Gene Runner software (version 3.02; Hastings Software, Inc.) and synthesized by Macrogene (South Korea). PCR products were visualized on prestained 1.5 % agarose gel after electrophoresis. For new alternative splice variants target PCR product fragments were extracted from Agarose gel by GeneJET gel extraction kit (Thermo Fisher Scientific, Inc) and sequenced directly (Bioneer, South Korea).

Quantitative measurement of gene expression was carried out in Bioer LineGene K and HotTaq EvaGreen qPCR q-PCR master mix (Cinnaclon, Iran) using the cycling condition of 15 min initial denaturation and enzyme activation at 95 °C, 40 cycles of 10 s denaturing at 95 °C, 15 s of annealing at 58 °C and 30 s of extending at 72 °C. Final melting curve analysis and gel electrophoresis were done to further confirm the PCR products accuracy.

Statistics

All measurements were performed in triplicate and two no-sample and no-RT negative controls were included. The Ct values were normalized to *GAPDH* as a house keeping gene and fold changes in gene expressions were normalized to HeLa cell line ($2^{-\Delta\Delta Ct}$). GraphPad Prism version 6.01 was used to analyze data statistically. One sample student's t test (fixed value = 1) was used for cell line gene expression experiment analysis and All data presented as log 10 of mean ± SE. For NT2 differentiation statistics, the 2-way ANOVA multiple comparison test (Bonferroni's) was carried out versus time point zero as a control. To explore any relation between gene expression levels of target genes a Pearson r correlation test was applied.

Results

Identification of a new format in alternative splicing of SOX2OT transcripts

Using specific primers (Table 1), SOX2OT transcript variants were amplified separately using RT-PCR in U-87 MG cell line. As presented in Fig. 2, while specific amplifications of variant 5 (F2, R4) and variant 6 (F2, R5) resulted in two fragments of 398 (Fig. 2a lanes 7–9) and 111 base pairs (Fig. 2a lanes 10–12); no expression of variants 2 and 3 were detected (data not shown). Surprisingly, gel electrophoresis indicated two extra amplified fragments when SOX2OT transcript 1 (F1, R1) and SOX2OT transcript 4 (F2, R1) specific primers were used for PCR reaction (Fig. 2a, lanes 2, 3 and 4–6). To examine nonspecific hybridization of primers, annealing temperature was increased to 3 °C in a gradient PCR (56, 57 and 58 °C); and the results indicated more intense amplification of same extra PCR fragments in both S1 and S4 reactions in high temperature conditions (Fig. 2a, compare lanes 1–3 and 4–6). Since both extra bands (∼168 and 129 bps) differ in length from original SOX2OT transcript variants 1 and 4 PCR products (454 and 415 bps) by about 286 nucleotides (equal to length of exon 8); we proposed existence of two new alternative splice variants lacking exon 8. To explore the idea, the 129 and 168 bps fragments were sequenced directly following agarose gel purification. Nucleic acid sequences revealed a new exon 7–9 junction in both proximal and distal SOX2OT transcript variants submitted as SOX2OT transcript variant 7 (GenBank: KF981435) and SOX2OT transcript variant 8 (GenBank:

Table 1 Specific primer sequences used for amplification of target genes

Primer	Primer sequence 5′→3′	Specific for
F1	TTCAGAAACTTATCAAGAGGTTC	SOX2OT exon 5 (variants 1, 2, 3, 8)
F2	TCTGTTCAGTATTTGGAAGAAAG	SOX2OT exon 6 (variants 4, 5, 6, 7)
R1	GCAGAGATTATCCGATTTGG	SOX2OT exon 9 (variants 1, 4, 7, 8)
R2	TTGGACCCGCGATGTC	SOX2OT exon 9–10 boundary (variants 1, 4)
R4	ACATTATTTCTAAGTTGGATATGTC	SOX2OT exon 8–10 boundary (variants 2, 5)
R5	CATTATTTCTAAGTTGGATTGGA	*SOX2OT* exon 7–10 boundary (variants 3, 6)
R6	GCTTGGACCCGCGTG	SOX2OT exon 7–9 boundary (variants 7, 8)
F3	TACAGCATGTCCTACTCGCAG	*SOX2*
R8	GAGGAAGAGGTAACCACAGGG	
F4	GAGCGAGATCCCTCCAAAAT	GAPDH
R9	GGCTGTTGTCATACTTCTCATG	
F5	TTCGCAAGCCCTCATTTCAC	*OCT4A*
R10	CGAGAAGGCGAAATCCGAAG	

KJ415055) respectively (Fig. 1c). Then for more assurance, a reverse primer spanning 7–9 exon boundary (R6) was used to amplify SOX2OT transcript variant 7 (104 bps, Fig. 2b, lane 15) and SOX2OT transcript variant 8 (139 bps, Fig. 2b, lane 13) specifically by RT-PCR. Similarly a reverse primer specific for the 9–10 exon junction was used for specific amplification of SOX2OT-1 (424 bps, Fig. 2b, lane 14) and SOX2OT-4 (389 bps, Fig. 2b, lane 16) in human cancer cell lines.

SOX2OT variant 7 and 8 are highly expressed in human cancer cell lines

We screened the expression level of two novel transcripts, SOX2OT-7 and SOX2OT-8 in five selected human cancer cell lines (HeLa, MCF7, HepG2, U-87 MG and NT2) by Real time RT-PCR. To facilitate the comparison of cell lines, HeLa cell line (which showed the least expression values) was considered as calibrator. Unexpectedly, all tested cancer cell lines, especially NT2 showed high levels of SOX2OT-7 and SOX2OT-8 expressions; except Mcf-7 which expresses the minimum level of SOX2OT-8 (Fig. 3a).

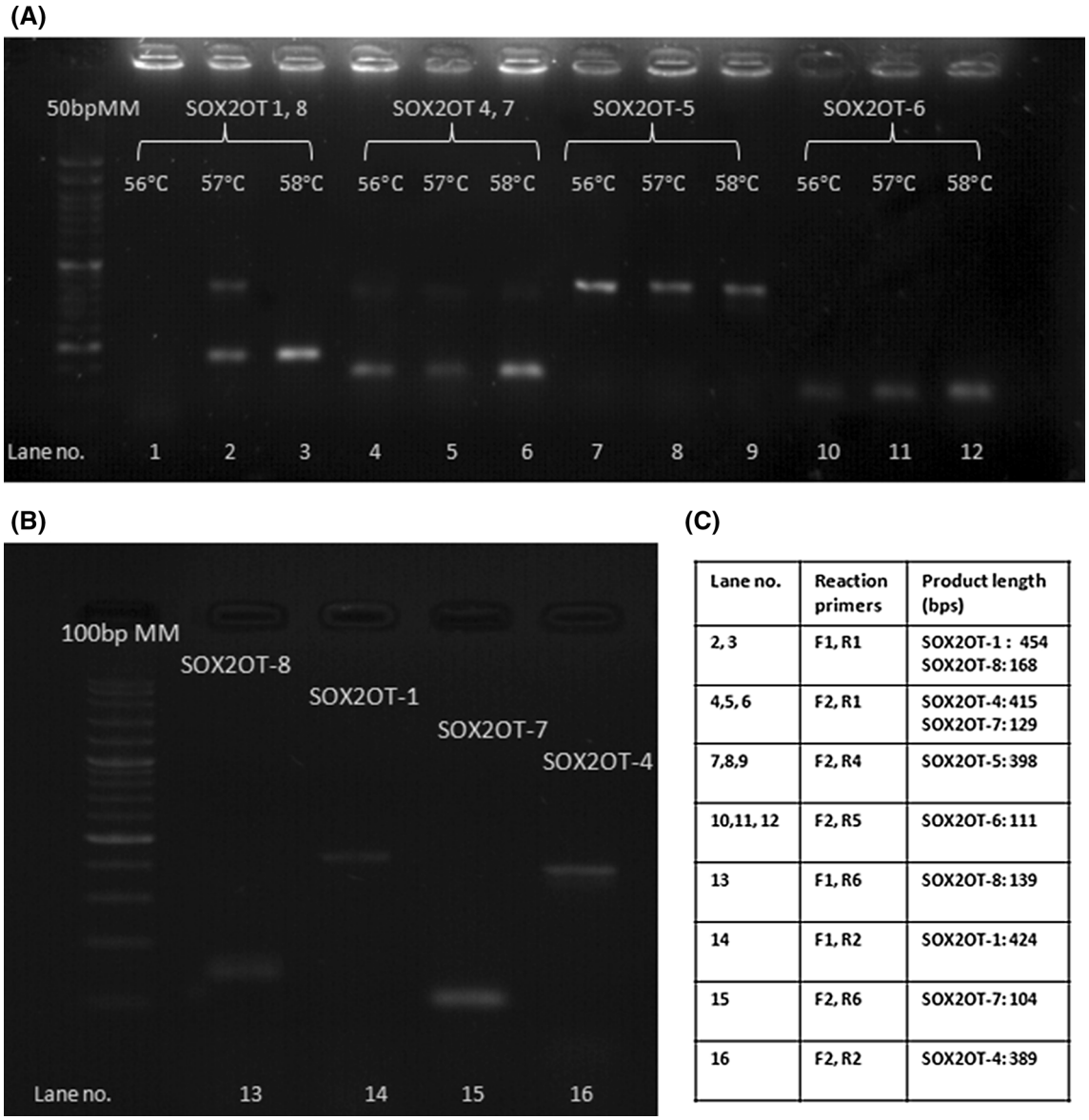
Regarding to the supposed association of SOX2OT transcripts and *SOX2* gene, we explored the expression levels of *SOX2* and *OCT4A*, the pluripotency regulators in accordance to SOX2OT-7 and SOX2OT-8. *SOX2* and *OCT4A* gene expressions were also quantified in all cell lines using the same method. As shown in Fig. 3a, high levels of *SOX2* expression was detected in all cell lines same as SOX2OT transcript variants; however *OCT4A* expression was observed in some of studied cell lines. One-sample student's *t* test (CI 95 %, alpha = 0.05) indicated that the means of *SOX2* (*P* value = 0.008), SOX2OT-7 (*P* value = 0.02), SOX2OT-8 (*P* value = 0.03) fold changes are significantly different from fixed theoretical value of 1. Pearson r correlation test (CI 90 %, alpha = 0.1) indicated significant correlation between expression values of SOX2OT-7 and SOX2OT-8 (r = 0.95, *P* value = 0.003), *SOX2* and SOX2OT-7 (r = 0.75, *P* value = 0.08).

SOX2OT-7 is the most abundant isoform of SOX2OT transcripts in NT2, U-87MG and HepG2 cancer cell lines

Considering the importance of SOX2OT-7 and SOX2OT-8 along with other SOX2OT variants, we studied the SOX2OT profile in three cancer cell lines (U-87MG, NT2, HepG2). The expression levels of SOX2OT transcript variants 1, 4, 5, 6, 7, 8 were specifically measured in triplicate by real time RT-PCR. The Ct values were normalized to GAPDH as an internal control and 1/ΔCt percent was illustrated to compare SOX2OT transcript variants in each cell line. Results showed that the transcript variants of SOX2OT 4, 5, 6, 7 and 8 are expressed in three tested cell lines and interestingly, SOX2OT-7 assign high expression scores in comparison to other SOX2OT transcripts (Fig. 3b). SOX2OT transcript variant 7 occupies >90 % of total SOX2OT transcript variants measured in NT2 teratocarcinoma cell line, >55 % in U-87MG glioma cell line and >35 % in HepG2 hepatocellular cancer cell line. No significant levels of SOX2OT-1 and SOX2OT-5 were detectable in three selected cancer cell line.

SOX2OT-7 expression decreases along with NT2 cell line neuronal induction

In relation to other transcript variants of SOX2OT, high expression levels of SOX2OT-7 transcript in NT2 teratocarcinoma cell line highlights a potential association of SOX2OT-7 with stemness features of the cells. We then interested to measure whether SOX2OT-7 expression



Lane no.	Reaction primers	Product length (bps)
2, 3	F1, R1	SOX2OT-1 : 454 SOX2OT-8: 168
4,5, 6	F2, R1	SOX2OT-4: 415 SOX2OT-7: 129
7,8,9	F2, R4	SOX2OT-5: 398
10,11, 12	F2, R5	SOX2OT-6: 111
13	F1, R6	SOX2OT-8: 139
14	F1, R2	SOX2OT-1: 424
15	F2, R6	SOX2OT-7: 104
16	F2, R2	SOX2OT-4: 389

Fig. 2 Gel electrophoresis of RT-PCR products of SOX2OT transcript variants amplification in U-87 MG cell line. **a** Temperature gradient of annealing temperatures 56–58 °C was applied for primer annealing step of RT-PCR reactions. Specific RT-PCR products of a 454 base pairs SOX2OT-1 product (*lane 2*); a 415 base pairs SOX2OT-4 product (*lanes 4–6*); a 398 product of SOX2OT-5 (*lanes 7–9*) and 111 base pairs SOX2OT-6 products (*lanes 10–12*) are shown. Noticeably extra bands of 168 bps in SOX2OT-1 reaction (*lanes 2–3*) and 129 bps in SOX2OT-4 reaction (*lanes 4–6*) have got more intensity parallel to annealing temperature increase. The length of extra fragments, interestingly differ about 286 nucleotides (exon 8) from original variants and potentially have been proposed as two new variants. **b** Using R2 and R6 Reverse primer specific for new exon boundaries, the specific amplification products of a 139 base pairs predicted SOX2OT-8 (*lane 13*), a 424 base pairs of SOX2OT-1 (*lane 14*), a 104 base pairs predicted SOX2OT-7 (*lane 15*) and a 389 base pairs of SOX2OT-4 (*lane 16*) are presented. **c** The primer and product description of each *lane* is presented in this table

specifically is altered during NT2 neuronal-like induction mediated by all-trans retinoic acid treatment. NT2 cells were induced to differentiate to neuronal-like lineage and cells were harvested for RNA extraction and gene expression analysis in six time points (0, 3, 7, 14, 21, 32 days). SOX2OT-7, *SOX2* and *OCT4A* gene expression levels were measured by Real time RT-PCR and normalized to GAPDH. Mean fold changes of gene expressions relative to Day zero are presented (Fig. 4). As expected, *SOX2* expression falls significantly from 7 to 32 days of differentiation (P value <0.01) and almost no *OCT4A* expression remains detectable after 7 days of differentiation (P value <0.0001). Notably, SOX2OT-7 expression declines upon NT2 differentiation similar to *SOX2* and *OCT4A*, two master regulators of pluripotency. Following 21 days of Retinoic acid treatment SOX2OT-7 expression

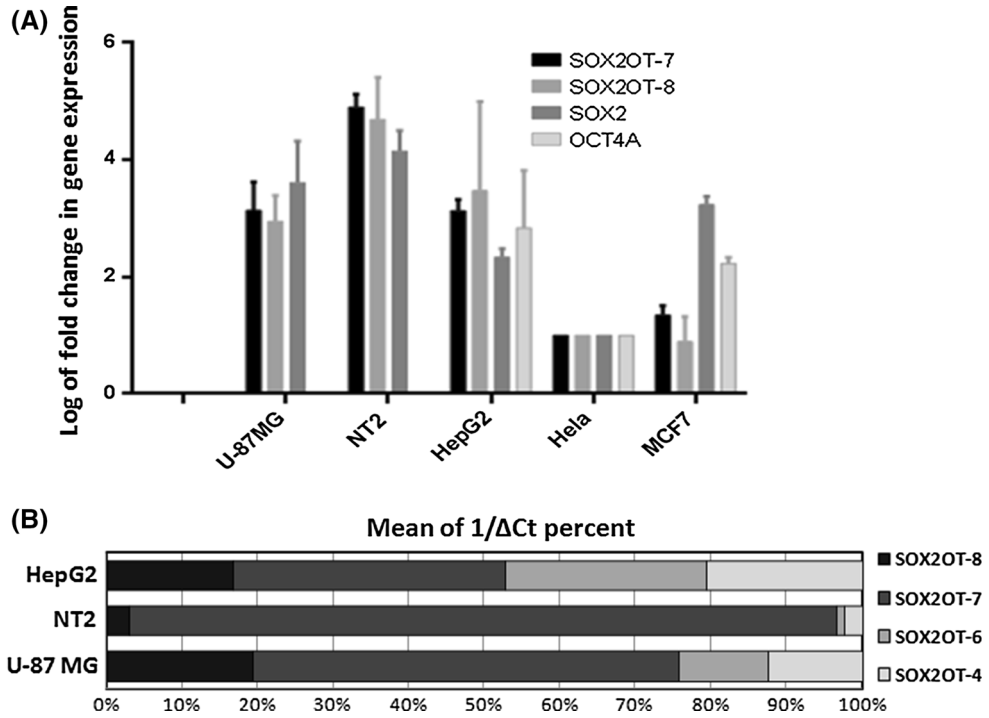


Fig. 3 Graph presentation of real time RT-PCR amplification of target genes. **a** Semi-quantitative comparison of SOX2OT-7, *SOX2OT-8*, *SOX2* and *OCT4A* genes expression level in five different human cancer cell lines. Log of fold change ($2^{-\Delta\Delta Ct}$) was computed relative to HeLa cell line. *GAPDH* gene expression was used as the housekeeping gene control to normalize Ct values. Notice the high expression of SOX2OT-7 and *SOX2OT-8* coincident with *SOX2* and *OCT4A* genes. **b** SOX2OT variants expressional profile of each cell line indicated as Mean of 1/ΔCt of each SOX2OT variants to total of all detectable variants × 100 in each cell line. High expressional score of SOX2OT-7 is remarkably illustrated in SOX2OT profile of each cell line

reduced up to zero (*P* value <0.0001); however it increased to 0.5 fold at 32th day.

Discussion

According to the fact that the primary sequence of a polynucleotide determines its three-dimensional structure, it is not amazing that alternative spliced or alternative promoter mediated of non-coding transcripts can potentially exert different features in stability, 3-D folding, localization and even function. The minimum free energy prediction of two SOX2OT-4 and SOX2OT-7 isoforms was calculated by RNA Fold web server (http://rna.tbi.univie.ac.at/cgi-bin/RNAfold.cgi). Surprisingly, we found that splicing of the exon 8 (286 nucleotides in length) can obviously affect the infrastructure of SOX2OT-4 transcript and increase the minimum free energy of SOX2OT-7 secondary structure about 110 kcal/mol (data not shown). Beside less predicted structural stability relative to variant 4, SOX2OT-7 has significantly more cellular concentration in NT2, HepG2 and U-87MG cell lines. It seems that the higher expression of SOX2OT-7 is regulated post-transcriptionally, since both SOX2OT-4 and SOX2OT-7 are transcribed from same promoter. Indeed, partial conformational entropy of SOX2OT-7 may be considered as a functional advantage providing the structural dynamic necessity for a functional non coding RNA.

Cell type specific regulation of a non-coding RNA expression provides a clue for its functional importance [22]. Recently, Askarian-Amiri et al. [19] demonstrated *SOX2OT* gene expression differentially in estrogen positive breast cancer tissue and cell lines in concordant with *SOX2*. They also showed that the overexpression of *SOX2OT* is directly associated with *SOX2* gene up-regulation. Hou et al. [20] ascertained the role of SOX2OT in lung cancer progression and illustrated SOX2OT overexpression in lung tumor tissue and cell line. Shahryari et al. [16] reported three human SOX2OT splice variants named SOX2OT, SOX2OT-S1 and SOX2OT-S2 (GenBanks: NR_004053.2, JN711430.1 and JN882275.1) later updated to SOX2OT transcript variants 4, 5 and 6 which are over expressed in esophagus cancer tissues.

In the present study, instead of SOX2OT general gene expression measurement, we focused on different transcript variants of SOX2OT specifically in human cancer cell lines and identified two new SOX2OT transcript variants for the first time. We presented different SOX2OT transcripts that

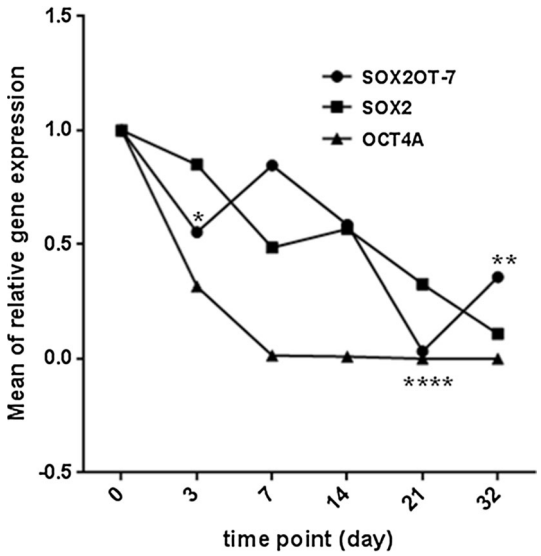


Fig. 4 Periodic changes of SOX2OT-7, *SOX2* and *OCT4A* expression during the neural differentiation of NT2 cells. NT2 cells were induced to Neuronal like lineage by ATRA treatment. Total RNA of three biological replicates at 0, 3, 7, 14, 21 and 32 days after treatment were then extracted and used for gene expression analysis by qRT-PCR ($2^{-\Delta\Delta Ct}$). All time points are presented relative to time 0. *GAPDH* used as house-keeping gene. Notice the gradual decrease of all three genes although a slight increase observed in SOX2OT-7 at the end of treatment. (*P value = 0.04; **P value = 0.003; ****P value <0.0001)

among them two novel SOX2OT-7 and SOX2OT-8 are highly expressed in five different human cancer cell lines; suggesting the potential role of SOX2OT transcripts in carcinogenesis.

There are evidences indicating the probable role of SOX2OT in development and stem cell regulation. Different isoforms of sox2ot are precisely regulated during mammalian embryogenesis and like *SOX2*; sox2ot expression is down-regulated along with the differentiation of mouse ESC [17]. Breast cancer stem cell enrichment mediated by suspension culture increases SOX2OT and *SOX2* gene expression levels [19]. Indeed SOX2OT (variants 4, 5, 6) expression decrease in parallel with NT2 neuronal-like differentiation as same as *SOX2* and *OCT4A* [16]. Interestingly, in NT2 cancer cell line with stem cell like properties, high expression of SOX2OT-7 in comparison to other variants, highlights it as the SOX2OT isoforms most related to pluripotency. Here, we also demonstrated that the expression of SOX2OT-7 specifically decreases along with NT2 differentiation as same as *SOX2* and *OCT4A*, but there is a slight raise in SOX2OT-7 expression observable at the end of differentiation (day 32). Previously, Amaral et al. [17] have shown that sox2ot expression remains high when mouse embryonic body differentiates to mesoderm lineage.

It has been postulated that distal promoter of SOX2OT is merely active in neuronal lineage [17, 23]; however our results illustrated that both proximal and distal SOX2OT promoters are constitutively active in all studied cancer cell lines of different histological origins; which could be relevant to changes obtained during carcinogenesis. Although the functional properties of different SOX2OT isoforms are not exactly determined yet, but studies suggest that the association of SOX2OT with *SOX2* regulation in stem cells [16, 17, 19, 24] and G2/M cell cycle progression in cancer cells [16, 19]. Concerning the regulated expression and splicing of different transcript isoforms of SOX2OT as a long non-coding RNA, we propose that each SOX2OT may potentially function diversely and should be addressed in separate.

Conclusion

Our results suggest that the two novel alternative spliced variants of *SOX2* overlapping transcripts, SOX2OT-7 and SOX2OT- 8 are highly associated with different human cancer cell lines and are over expressed coinciding with *SOX2* gene which is transcribed in same direction. In comparison to other SOX2OT transcripts, cancer cell lines favor SOX2OT-7 and SOX2OT-8 transcript expression propounding a plausible function in carcinogenesis. The marked direct association of SOX2OT-7 and stemness features of NT2 cell line, indicates its potential role in pluripotency. Counting these two novel spliced variants, altogether eight different transcripts of SOX2OT have been identified to date and their functional differences still remain to be investigated.

Acknowledgments This study was part of a PhD thesis financially supported by Golestan University of Medical Science (grant number: 92041130).

References

1. Birney E, Stamatoyannopoulos JA, Dutta A, Guigó R, Gingeras TR, Margulies EH, Weng Z, Snyder M, Dermitzakis ET, Thurman RE (2007) Identification and analysis of functional elements in 1 % of the human genome by the ENCODE pilot project. Nature 447(7146):799–816
2. Carninci P, Kasukawa T, Katayama S, Gough J, Frith M, Maeda N, Oyama R, Ravasi T, Lenhard B, Wells C (2005) The transcriptional landscape of the mammalian genome. Science 309(5740):1559–1563
3. Consortium EP (2004) The ENCODE (ENCyclopedia of DNA elements) project. Science 306(5696):636–640
4. Consortium IHGS (2004) Finishing the euchromatic sequence of the human genome. Nature 431(7011):931–945
5. Amaral PP, Mattick JS (2008) Noncoding RNA in development. Mamm Genome 19(7–8):454–492
6. Kapranov P, Cheng J, Dike S, Nix DA, Duttagupta R, Willingham AT, Stadler PF, Hertel J, Hackermüller J, Hofacker IL

(2007) RNA maps reveal new RNA classes and a possible function for pervasive transcription. Science 316(5830): 1484–1488
7. Ponting CP, Oliver PL, Reik W (2009) Evolution and functions of long noncoding RNAs. Cell 136(4):629–641
8. Chen L-L, Carmichael GG (2010) Decoding the function of nuclear long non-coding RNAs. Curr Opin Cell Biol 22(3):357–364
9. Kaikkonen MU, Lam MT, Glass CK (2011) Non-coding RNAs as regulators of gene expression and epigenetics. Cardiovasc Res 90(3):430–440
10. Taft RJ, Pang KC, Mercer TR, Dinger M, Mattick JS (2010) Non-coding RNAs: regulators of disease. J Pathol 220(2):126–139
11. Wang X, Song X, Glass CK, Rosenfeld MG (2011) The long arm of long noncoding RNAs: roles as sensors regulating gene transcriptional programs. Cold Spring Harb Perspect Biol 3(1):a003756
12. Wilusz JE, Sunwoo H, Spector DL (2009) Long noncoding RNAs: functional surprises from the RNA world. Genes Dev 23(13):1494–1504
13. Fejes-Toth K, Sotirova V, Sachidanandam R, Assaf G, Hannon GJ, Kapranov P, Foissac S, Willingham AT, Duttagupta R, Dumais E (2009) Post-transcriptional processing generates a diversity of 5′-modified long and short RNAs. Nature 457(7232):1028–1032
14. Fantes J, Ragge NK, Lynch S-A, McGill NI, Collin JRO, Howard-Peebles PN, Hayward C, Vivian AJ, Williamson K, van Heyningen V (2003) Mutations in SOX2 cause anophthalmia. Nat Genet 33(4):461–463
15. Andrew T, Maniatis N, Carbonaro F, Liew SM, Lau W, Spector TD, Hammond CJ (2008) Identification and replication of three novel myopia common susceptibility gene loci on chromosome 3q26 using linkage and linkage disequilibrium mapping. PLoS Genet 4(10):e1000220
16. Shahryari A, Rafiee MR, Fouani Y, Oliae NA, Samaei NM, Shafiee M, Semnani S, Vasei M, Mowla SJ (2014) Two novel splice variants of SOX2OT, SOX2OT-S1, and SOX2OT-S2 are coupregulated with SOX2 and OCT4 in esophageal squamous cell carcinoma. Stem Cells 32(1):126–134
17. Amaral PP, Neyt C, Wilkins SJ, Askarian-Amiri ME, Sunkin SM, Perkins AC, Mattick JS (2009) Complex architecture and regulated expression of the Sox2ot locus during vertebrate development. RNA 15(11):2013–2027
18. Hussenet T, Dali S, Exinger J, Monga B, Jost B, Dembelé D, Martinet N, Thibault C, Huelsken J, Brambilla E (2010) SOX2 is an oncogene activated by recurrent 3q26.3 amplifications in human lung squamous cell carcinomas. PLoS One 5(1):e8960
19. Askarian-Amiri ME, Seyfoddin V, Smart CE, Wang J, Kim JE, Hansji H, Baguley BC, Finlay GJ, Leung EY (2014) Emerging role of long non-coding RNA SOX2OT in SOX2 regulation in breast cancer. PLoS One 9(7):e102140
20. Hou Z, Zhao W, Zhou J, Shen L, Zhan P, Xu C, Chang C, Bi H, Zou J, Yao X (2014) A long noncoding RNA Sox2ot regulates lung cancer cell proliferation and is a prognostic indicator of poor survival. Int J Biochem Cell Biol 53:380–388
21. Atlasi Y, Mowla SJ, Ziaee SA, Gokhale PJ, Andrews PW (2008) OCT4 spliced variants are differentially expressed in human pluripotent and nonpluripotent cells. Stem Cells 26(12):3068–3074
22. Mattick JS, Makunin IV (2006) Non-coding RNA. Hum Mol Genet 15(suppl 1):R17–R29
23. Mercer TR, Dinger ME, Sunkin SM, Mehler MF, Mattick JS (2008) Specific expression of long noncoding RNAs in the mouse brain. Proc Natl Acad Sci 105(2):716–721
24. Shahryari A, Jazi MS, Samaei NM, Mowla SJ (2015) Long non-coding RNA SOX2OT: expression signature, splicing patterns, and emerging roles in pluripotency and tumorigenesis. Front Genet 6:196