King-Hwa Ling, Peter J. Brautigan, Sarah Moore, Rachel Fraser, Pike-See Cheah, Joy M. Raison, Milena Babic. Young Kyung Lee, Tasman Daish, Deidre M. Mattiske, Jeffrey R. Mann, David L. Adelson, Paul Q. Thomas, Christopher N. Hahn, Hamish S. Scott

Derivation of an endogenous small RNA from double-stranded Sox4 sense and natural antisense transcripts in the mouse brain

Genomics, 2016; 107(2-3):88-99

Originally published at:
http://doi.org/10.1016/j.ygeno.2016.01.006

Genomics 107 (2016) 88–99

Contents lists available at ScienceDirect

Genomics

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

Derivation of an endogenous small RNA from double-stranded *Sox4* sense and natural antisense transcripts in the mouse brain

King-Hwa Ling [a,b,c,*], Peter J. Brautigan [a], Sarah Moore [a], Rachel Fraser [a], Pike-See Cheah [c,d,e], Joy M. Raison [f], Milena Babic [a], Young Kyung Lee [a], Tasman Daish [d], Deidre M. Mattiske [g], Jeffrey R. Mann [h], David L. Adelson [d], Paul Q. Thomas [d], Christopher N. Hahn [a], Hamish S. Scott [a,b,**]

[a] Department of Molecular Pathology, The Institute of Medical and Veterinary Science and The Hanson Institute, P.O. Box 14 Rundle Mall Post Office, Adelaide, SA 5000, Australia
[b] School of Medicine, Faculty of Health Sciences, University of Adelaide, Adelaide, SA 5005, Australia
[c] NeuroBiology & Genetics Group, Genetics and Regenerative Medicine Research Centre, Faculty of Medicine and Health Sciences, Universiti Putra Malaysia, 43400 Serdang, Selangor DE, Malaysia
[d] School of Molecular and Biomedical Science, Faculty of Sciences, University of Adelaide, Adelaide, SA 5005, Australia
[e] Department of Human Anatomy, Faculty of Medicine and Health Sciences, Universiti Putra Malaysia, 43400 UPM Serdang, Selangor DE, Malaysia
[f] eResearchSA, University of Adelaide, North Terrace, Adelaide, SA 5005, Australia
[g] Theme of Laboratory and Community Genetics, Murdoch Childrens Research Institute, Royal Children's Hospital, Flemington Road, Parkville, VIC 3052, Australia
[h] Biomedicine Discovery Institute, Monash University, VIC 3800, Australia

ARTICLE INFO

Article history:
Received 20 September 2015
Received in revised form 18 January 2016
Accepted 20 January 2016
Available online 21 January 2016

Keywords:
Noncoding RNA
Small regulatory RNA
piRNAs
Endo-siRNAs
Mouse brain development
Antisense RNA
Natural antisense transcripts

ABSTRACT

Natural antisense transcripts (NATs) are involved in cellular development and regulatory processes. Multiple NATs at the *Sox4* gene locus are spatiotemporally regulated throughout murine cerebral corticogenesis. In the study, we evaluated the potential functional role of *Sox4* NATs at *Sox4* gene locus. We demonstrated *Sox4* sense and NATs formed dsRNA aggregates in the cytoplasm of brain cells. Over expression of *Sox4* NATs in NIH/3T3 cells generally did not alter the level of *Sox4* mRNA expression or protein translation. Upregulation of a *Sox4* NAT known as *Sox4ot1* led to the production of a novel small RNA, *Sox4_sir3*. Its biogenesis is Dicer1-dependent and has characteristics resemble piRNA. Expression of *Sox4_sir3* was observed in the marginal and germinative zones of the developing and postnatal brains suggesting a potential role in regulating neurogenesis. We proposed that *Sox4* sense-NATs serve as Dicer1-dependent templates to produce a novel endo-siRNA- or piRNA-like *Sox4_sir3*.

1. Introduction

Less than 2% of the 3 billion bases in the human genome consist of protein coding sequences [1,2]. The Encyclopedia of DNA Elements (ENCODE) project showed as much as 90% of the studied 30 million bases of the human genome were transcribed, mostly into non-protein coding RNA [3]. This finding is in agreement with previously reported transcriptional landscapes of the mammalian genome [4–6] suggesting the previously termed 'junk' DNA contributes far more to the transcriptional profile of an organism. These non-protein coding transcripts differ from the canonical protein-coding transcripts and other structural RNAs (*e.g.* rRNA and tRNA) in terms of their gene distribution, abundance of regulatory motifs and transcription start sites, and evolutionary constraint [3,7]. As a consequence, these transcripts are difficult to characterise functionally.

One type of non-protein coding transcript has sequences that are partially/fully complementary to a protein-coding gene or RNA. These transcripts are known as natural antisense transcripts (NATs) and are either transcribed from the opposite DNA strand of the same protein-coding gene locus (known as *cis*-NATs) or from different loci within the genome (known as *trans*-NATs). To date, numerous NATs have been discovered in the mammalian genome and up to 20% of protein coding genes in human and mouse have at least one overlapping NAT [8–11]. Although numerous sense-NAT pairs have been identified *in silico* in the mammalian genome, only a handful of NATs such as *Evf-2* [12], *Air* [13], *HOTAIR* [14] and *Kcnq1ot1* [15] are implicated in a significant biological mechanism. These NATs are mainly nuclear localised and regulate gene expression during development through transcriptional activation, transcriptional repression or *via* chromatin modification. Differential regulation of NATs expression has been implicated in various human disorders such as fragile X-associated tremor and ataxia

* Correspondence to: K.-H. Ling, Genetics and Regenerative Medicine Research Centre, Faculty of Medicine and Health Sciences, Universiti Putra Malaysia, 43400 Serdang, Selangor DE, Malaysia.
** Correspondence to: H.S. Scott, School of Medicine, Faculty of Health Sciences, University of Adelaide, Adelaide, SA 5005, Australia.
E-mail addresses: lkh@upm.edu.my (K.-H. Ling), hamish.scott@health.sa.gov.au (H.S. Scott).

http://dx.doi.org/10.1016/j.ygeno.2016.01.006

syndrome [16], breast, renal and colon cancer [17–19], human follicular lymphoma [20], Beckwith–Wiedemann syndrome [21] and alpha-thalassemia [22], implicating the involvement of NATs in disease development. However, the mechanism underlying the role of NATs in disease progression remains poorly characterised.

We previously reported that the *Sox4* (Sry-related HMG box 4 gene) gene locus produced multiple overlapping protein-coding mRNAs and NATs during mouse cerebral corticogenesis [23]. *Sox4* is a single exon gene encoding a 47 kDa transcription factor containing a high-mobility group (HMG) domain, which functions in DNA binding, DNA bending, protein interactions and nuclear import or export [24,25]. Sox4 binds the DNA sequence (A/T)(A/T)CAA(A/T)G (or IUPAC code WWCAAW) and in the presence of one or more specific co-factors, can function as either a transcription repressor or activator [25,26] that regulates various developmental processes such as lymphocyte differentiation and bone, heart and brain development [27–31]. Its expression has also been implicated in the progression or transformation of various tumours and cancerous cells [32–37]. In the central nervous system, Sox4 regulates pan-neuronal gene expression that involves in the establishment of neuronal properties [30], and must be downregulated for proper myelination to occur [31]. The expression of multiple *cis*-NATs at the *Sox4* gene locus may be required for proper Sox4 protein synthesis through regulation of protein-coding transcripts at either the transcription, post-transcription or translation level.

Sense and antisense transcripts may form double-stranded RNAs (dsRNAs) or fold into conserved secondary hairpin structures that serve as templates for the biogenesis of small RNAs [38]. Small interfering RNA (siRNA), microRNA (miRNA) and PIWI-interacting RNA (piRNA) are 3 major categories of small RNAs. PiRNA is the largest class of small non-coding RNA molecules (24–32 nt) [39,40] that are distinct from miRNA (~22 nt) or siRNA (~20–25 nt) [41]. PiRNAs have been associated with epigenetic or post-transcription silencing of transposable elements in germ line cells [42,43] whereas both miRNAs and siRNAs target mainly mRNAs to exert translational repression or mRNA degradation [44,45]. The origin of small RNAs has been extensively characterised and located at both intergenic and intragenic regions [46,47] including tRNA genes [48,49]. Biogenesis of these small RNAs are different from each other and has been extensively reviewed previously [41,45].

To investigate the role of *Sox4 cis*-NATs during brain development, we characterised NATs at the *Sox4* gene locus, and their underlying regulatory effect on the *Sox4* protein-coding transcripts. In this study, we demonstrated that *Sox4 cis*-NATs is capable of forming double stranded RNA with the *Sox4* protein-coding (sense) transcripts in the brain. Upregulation of a novel *Sox4* NATs known as *Sox4ot1* in an artificial model lead to increased production of a novel small RNA with unconfirmed siRNA or piRNA characteristics. The small RNA was found expressed specifically in the germinative layers of the mouse brain suggesting a rare phenomenon in the mammalian system where sense-NATs serve as Dicer1-dependent templates in the cytoplasm to produce a siRNA- or piRNA-like small RNA.

2. Materials and methods

2.1. Breeding and handling of animals

Breeding and handling of animals were carried out according to guidelines approved by the Melbourne Health Animal Ethics Committee (Project numbers 2001.045 and 2004.041) and the University of Adelaide Animal Ethics Committee (S-086-2005). All mice used in the study were C57BL/6 unless otherwise specified. Mice were kept under a 12 h light/12 h dark cycle with unlimited access to food and water and were sacrificed by CO_2 asphyxiation. Brain tissues and mouse organs were dissected as described previously [23].

2.2. Rapid amplification of cDNA ends (RACE)-southern analysis

Total RNA was extracted from the E15.5 cerebral cortex using RNeasy Lipid Tissue Mini Kit (QIAGEN) according to the manufacturer's protocol. One thousand nanogram of pooled total RNA ($n = 3$) was used for 5′ and 3′ RACE analyses of *Sox4* NATs. 5′ and 3′ RACE were carried out using SMART™ RACE Amplification Kit (Clonetech) according to the manufacturer's protocol. For both 5′ and 3′ RACE, up to seven *Sox4* specific primers were included for first strand cDNA synthesis. These primers were located over the ~5 kb *Sox4* gene locus. Following cDNA synthesis, PCR was individually performed using all seven primers and a 5′ universal primer for 5′ RACE or 3′ adaptor primer for 3′ RACE (Clontech). Amplified products were blotted onto Hybond-N+ nylon membrane (GE Healthcare) and probed using independent oligonucleotides designed across the ~5 kb *Sox4* gene locus to determine the specific *Sox4* amplicons. Independent oligonucleotide probes were end-labelled using [γ-^{32}P]-dATP and T4 polynucleotide kinase (Promega) according to the manufacturer's protocol. Five and six probes were used in 5′ and 3′ RACE analysis, respectively. See Supplementary GenBank File in [50] for primers and probe sequences; it is recommended that the file is visualised using Artemis Genome Browser and Annotation Tool [51]. Pre-hybridisation, hybridisation, washing and detection procedures were performed according to previously published protocols [23,52].

2.3. Real-time quantitative polymerase chain reaction (RT-qPCR)

Total RNA was isolated from various mouse organs and brain regions using TRIzol® Reagent (Invitrogen). Contaminating genomic DNA (gDNA) was removed from the total RNAs using the DNA-*free*™ Kit (Applied Biosystems) according to the manufacturers' protocols. Reverse transcription was performed using the Superscript® III Reverse Transcriptase Kit (Invitrogen) according to the manufacturer's protocol. RT-qPCR and relative quantification analyses were carried out according to published methods [23,52]. The stemloop RT-qPCR analysis of *Sox4_sir3* was performed based on a published protocol [53] by using specific primers such as stem loop primer (5′-GTTGGCTCT GGTAGGATG CCGCTCTCA GGGCATCCT ACCAGAGCCA AACGGAATC-3′, GeneWorks), a universal reverse primer (5′-GTAGGATGCC GCTCTCAGG-3′, GeneWorks) and a specific forward primer for *Sox4_sir3* (5′-TCTGACTCAA GGACAGCG AC AA-3′, GeneWorks).

In all relative quantification analysis, One-way Analysis of Variance (ANOVA) was used to compare the expression levels among groups, brain tissues or mouse organs. A P value of <0.05 was considered statistically significant. The least significant difference(s) (LSD) were provided when significant differences were detected among the groups.

2.4. RNA fluorescence in situ hybridisation (RNA FISH)

Cells from the 56-day old (P56) adult mouse cerebral cortex, hippocampus, olfactory bulbs and cerebellum were obtained, fixed and digested with RNase A. These procedures and the preparation of non-overlapping complementary RNA (cRNA) probes for *Sox4* sense and antisense transcripts (see Supplementary GenBank File in [50]), hybridisation, washing steps, staining and mounting of slides were carried out essentially as described previously [52]. Both sense and antisense cRNAs were labelled using two different Universal Linkage System (ULS) haptens/dyes: Dynomics547-ULS (absorbance at 547 nm and emission at 565 nm) for sense cRNA probe and Dynomics415-ULS (absorbance at 415 nm and emission at 472 nm) for antisense cRNA probe. The acquisition of immunofluorescence cell images was performed by using a Zeiss Axioplan 2 Imaging upright microscope with Axiovision software. Images were acquired when slides were observed under the 100X objective lens with oil immersion.

2.5. Locked nucleic acid-in situ hybridisation (LNA-ISH)

Preparation of paraffin embedded sections (8 μm), pre-hybridisation, hybridisation, washing, antibody reaction and colour development steps were performed essentially as described previously [53]. Custom-made *Sox4_sir3* LNA probes (Cat. no: EQ-70537, Exiqon) were used for the LNA-ISH experiment. Whole brain images were acquired by using an Olympus DP70 digital camera mounted on a Nikon SMZ1000 dissection microscope with AnalySIS software.

2.6. Mouse embryonic stem (mES) cells with $Dicer1^{C}$ allele

Mouse embryonic stem (mES) cells with DICER1 activity were obtained from a line heterozygous for a conditionally mutant *Dicer1* allele ($Dicer1^{c}$) and a null *Dicer1* allele ($Dicer1^{-}$). These genetic modifications were described previously [55]. mES cells without Dicer1 activity were produced by transient transfection of this $Dicer1^{c/-}$ line with Cre recombinase to produce $Dicer1^{-/-}$ subclones [82].

2.7. NIH/3T3 mouse fibroblast cell line

NIH/3T3 cells were obtained from American Type Culture Collection (www.atcc.org) and maintained in Dulbecco's Modified Eagle's Medium (Sigma Aldrich) supplemented with 10% (*v*/v) heat-inactivated foetal calf serum (FCS; Invitrogen), 100 units/ml penicillin, 100 μg/ml streptomycin and 2 mM L-glutamine. Cells were subcultured into new dishes when they reached 80% confluence or less using approximately 3–5×10^3 cells/cm^2 inoculums.

2.8. Overexpression of Sox4 NATs

Full length *Sox4* NATs were amplified from mouse gDNA using the paired-end ditags sequences as primers (see Supplementary GenBank File in [50]). Proofreading polymerase enzyme from Expand Long Template PCR System Kit (Roche Diagnostics) was used to amplify *Sox4* NATs (with sizes between 0.8 and 3 kb) from gDNA. Amplicons were blunt-end-cloned into pcDNA3 vector (Invitrogen) at EcoRV restriction site. The right clones were screened by orientation- and *Sox4*-specific PCR. The sizes of the constructs were validated by gel electrophoresis. Confirmed *Sox4* NAT constructs, namely, PET2, PET3, PET5 and PET6, were individually co-transfected into NIH/3T3 cells together with pcDNA3-eGFP constructs using Lipofectamine™ 2000 Transfection Reagent (Invitrogen) according to the manufacturer's protocol. PET construct to pcDNA-eGFP construct ratio used during transfection was 3:2. Equal number of molecules for PET2, PET3, PET5 and PET6 were considered in each transfection and pUC18 vector was added to normalise the total amount of DNA transfected (6 μg per 2 million cells) into NIH/3T3 cells. Each transfection for PET construct was controlled by a mock transfection using pcDNA3-empty vector and without any DNA (blank control).

2.9. Immunoblotting analysis

Approximately 24 h after transfection, cells were lysed in ice-cold lysis buffer [(50 mM Tris.HCl, pH 7.4, with 1% NP-40, 150 mM NaCl, 2 mM ethylene glycol-bis(beta-aminoethyl ether)-N,N,N,N-tetraacetic acid, 1 mM $NaVO_4$, 100 mM NaF, 10 mM $Na_4P_2O_7$, and EDTA-free protease inhibitor cocktail (Roche)]. Protein concentrations were assayed using Bradford Reagent (BioRad) according to the manufacturer's protocol. Equal amounts of protein (~20 μg) were loaded onto 10% acrylamide gels, separated by sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE). Electrophoresed proteins were transferred onto Amersham Hybond™-P hydrophobic polyvinylidene difluoride (PVDF) membrane followed by blocking with 5% (*w*/*v*) skim milk powder with 0.1% (*v*/v) Triton-X100 prepared in 1X PBS. The membrane was probed with a primary polyclonal rabbit antibody directed against Sox4 (Cat. No.: AB80261, Abcam) or polyclonal goat antibody directed against actin (Cat. No.: SC-1616, Santa Cruz) at 4 °C overnight followed by ImmunoPure® Goat Anti-Rabbit IgG, (H + L) Peroxidase Conjugate (Cat. No.: 31460, Pierce) or Polyclonal Rabbit Anti-Goat Immunoglobulins/HRP (Cat. No.: P0449, Dako) secondary antibody. Reactive bands were detected using Amersham ECL Plus™ Western Blotting Detection Reagents (GE Healthcare) according to the manufacturer's protocol. Pixelation analyses of bands were performed using ImageJ software according to the standard protocol published at http://rsb.info.nih.gov/ij.

3. Results

3.1. Sox4 NATs have multiple transcription start sites and polyadenylation sites

To characterise *Sox4* NATs that were expressed during brain development, we performed 5′ and 3′ RACE-southern analyses using E15.5 mouse cerebral cortex total RNA and combinations of primers/probes designed across the *Sox4* gene locus (Fig. 1A–D). To aid identification of full length *Sox4* NATs, we mapped previously described *Sox4* SAGE tags generated from 4 developmental stages of the mouse cerebral cortex to both sense and antisense strands of the *Sox4* gene locus [23]. Additional information from the 3′ RACE analysis of *Sox4* NATs in the same study was also mapped to our data (Fig. 1D). Next, we compared our results to the FANTOM Paired-End Ditags (PET) sequences, which were obtained from the Ensembl website (www.ensembl.org) and mapped to the *Sox4* gene locus (Fig. 1E). All primers, probes and annotations at *Sox4* gene locus are provided in one GenBank file (see Supplementary GenBank File in [50]).

We identified multiple transcription start sites (TSSs) as well as polyadenylation sites for *Sox4* NATs across the entire *Sox4* gene locus. The 5′ RACE analysis showed 19 different TSSs for *Sox4* NATs with 9 potential TSSs based on the prominent bands in southern analysis (Fig. 1A and C). A total of 12 polyadenylation sites were found for *Sox4* NATs based on 3′ RACE analysis, with four of them represented by prominent bands in southern analysis (Fig. 1B and D). This suggests that transcripts with different TSSs or polyadenylation sites have different expression levels.

We compared our RACE results with the mapped PET sequences (Fig. 1E). PET1-5 have TSSs that corresponded well (within ±100 nt) to our mapped TSSs of *Sox4* NATs based on 5′ RACE analysis, whereas the TSS for PET6 was found >100 nt upstream to the TSS of our dataset. SAGE tags generated from the most 3′ regions of PolyA+ RNAs confirmed 8 of the mapped polyadenylation sites for *Sox4* NATs. Of these, 3 polyadenylation sites from the previous study [23] matched with PET3. Other polyadenylation sites without any SAGE tags (beyond L6 in Fig. 1D) also matched (within ±100 nt) with the polyadenylation sites for PET1 and PET6. We did not find any polyadenylation sites from our 3′ RACE analysis that matched the polyadenylation sites of PET2, PET4 and PET5. This result could be attributed to the primer and/or probe used in our RACE experiments, which determine the specificity or resolution of the analysis.

Next, we searched and mapped all the TATA box sequences, and 12 polyadenylation signal variants [56] across the antisense strand of the *Sox4* gene locus (Fig. 1E). Only six possible TATA boxes were found with two containing a very conserved TATA box sequence, TATAAA AAA. The TATA box was located upstream of the TSSs of *Sox4* NATs from our 5′ RACE analysis (between L1 and L3 in Fig. 1C) and PET2, PET3 and PET4. On the other hand, polyadenylation signal sequences were found in all the mapped polyadenylation sites of *Sox4* NATs and in all the PETs except for PET4 (>500 nt from polyA site) and PET5 (no signal). The analysis shows that these *Sox4* NATs utilised alternative core promoter sequences and polyadenylation signals during transcription initiation and polyadenylation, respectively.

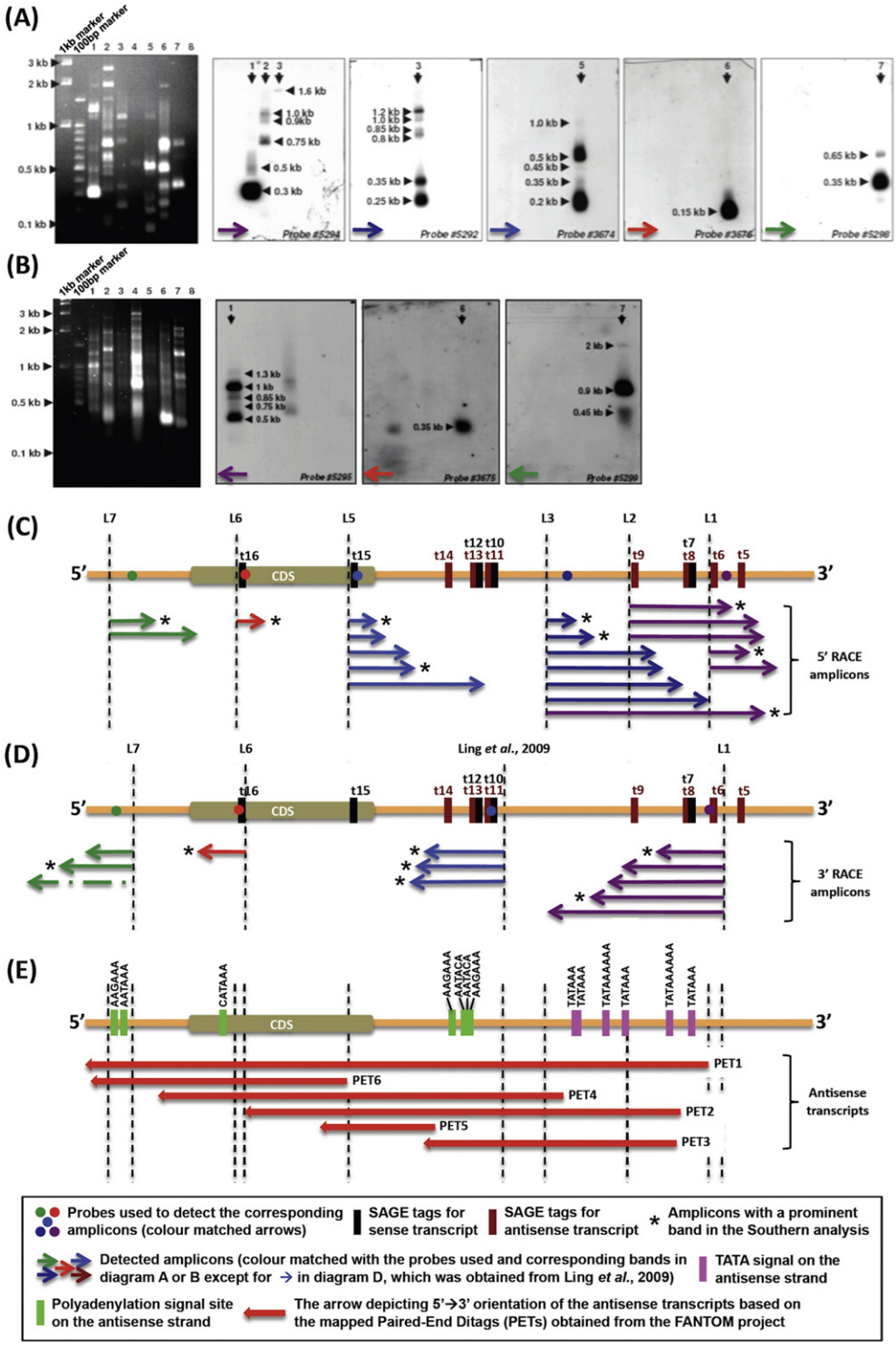

Fig. 1. RACE-southern analysis of *Sox4* antisense transcripts expressed in the E15.5 cerebral cortex. 5′ RACE for *Sox4* antisense transcripts was independently carried out using the universal primer for 5′ and each of the 7 gene specific primers designed across the *Sox4* gene locus (A). Amplicons for each reaction were electrophoresed, blotted and specific *Sox4* antisense amplicons were detected using independent oligonucleotide probes designed downstream to each of the original primer used. Similar approach was performed for 3′ RACE analysis (B). The oligonucleotide probes used for detection are given at the lower right corner of each gel photo and represented by a coloured arrow at the lower left corner. These colour arrows denote all the corresponding amplicons in each gel and are schematically represented in figures (C) and (D) for both 5′ and 3′ RACE analyses, respectively. FANTOM Paired Ends di-Tag (PET) sequences for *Sox4* antisense transcripts, which were obtained from Ensembl website (www.ensembl.org) were mapped to the *Sox4* gene locus (E). Previously reported SAGE tags (Ling et al., 2009), selected TATA box and polyadenylation signal sequences were also mapped to the gene locus. Detail legend descriptions are provided in the bottom panel of the figure.

3.2. Sox4 sense and NATs form cytoplasmic dsRNA aggregates in brain cells

To determine whether *Sox4* NATs are functional, we first evaluated the cellular localisation of these transcripts in relation to sense transcripts. First, we performed RNA FISH analysis using non-overlapping RNA probes for both sense and antisense transcripts on cells obtained from the adult cerebral cortex, hippocampus, olfactory bulbs and cerebellum (Fig. 2; see Supplementary GenBank File in [50]

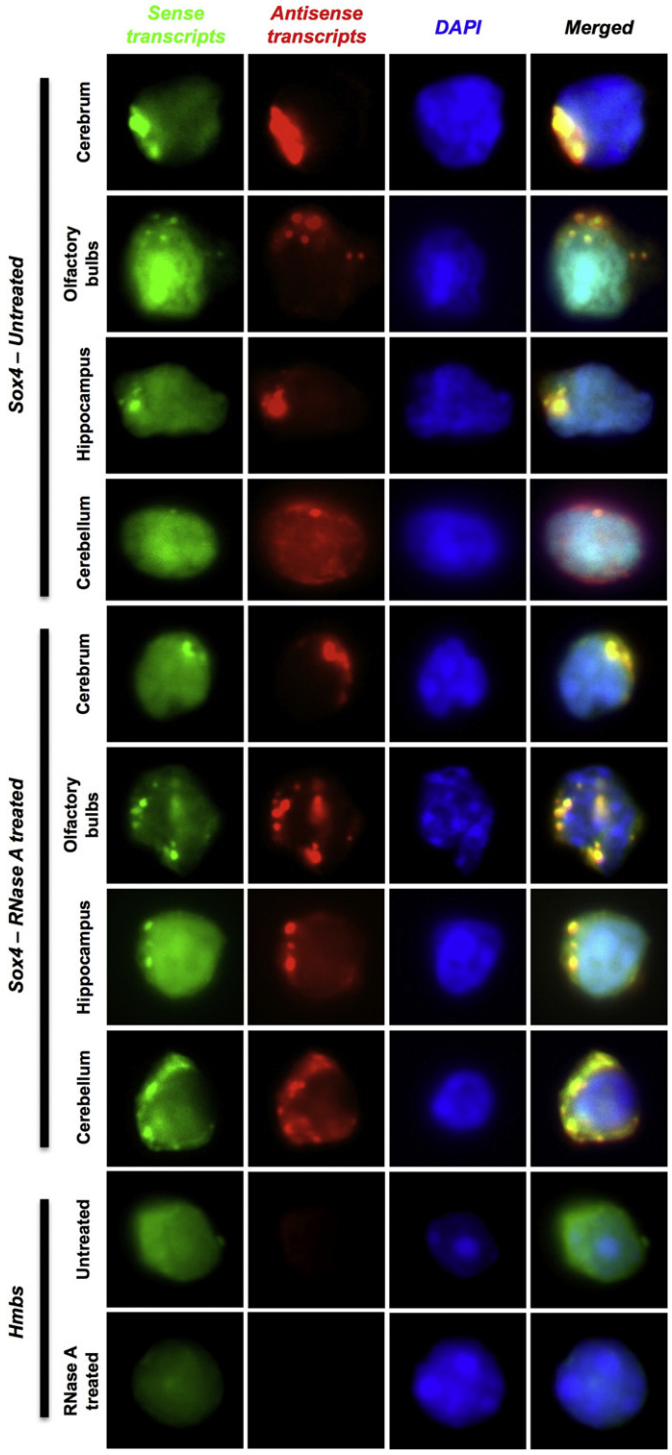

Fig. 2. RNA FISH of *Sox4* sense and NATs. RNA FISH of *Sox4* sense and NATs was performed on trypsinised cells obtained from different regions of the adult mouse brain. The type of transcripts analysed is shown at the top of the figure and the origin of cells is shown to the left of the micrographs. Scale bars = 5 μm.

for probes information). The analysis showed cytoplasmic co-localisation of sense and NATs as aggregates in 5–10% of cells assessed. No sense and NATs aggregates or signals were found in the nucleus suggesting *Sox4* NATs are not playing a direct role in regulating *Sox4* gene transcription or chromatin modification. Next, we asked whether NATs formed double stranded RNA with the sense transcript in the cytoplasm, an event that is important in RNA interference and translation regulation. We performed RNA FISH on RNase A treated cells from the same mouse brain regions (Fig. 2). Additional RNA FISH experiments targeting the *Hmbs* housekeeping gene were performed as a control for the RNase A digestion step (see Fig. 2 in [50]). Treatment of cells with RNase A did not cause any change in RNA FISH, compared to non-treated cells (Fig. 2). This is an indirect suggestion that *Sox4* NATs may interact with Sox4 sense transcripts and form dsRNA in the cytoplasm. The z-stacking analysis of these cells is provided in movie file format (see Supplementary Movies in [50]).

3.3. Sox4 NATs did not affect the transcription or translation of Sox4 mRNA in NIH/3T3 cells

To determine whether *Sox4* NATs affect the translation of Sox4 protein, we overexpressed selected PET2 (3.216 kb), PET3 (1.919 kb), PET5 (0.807 kb) and PET6 (1.824 kb) NATs in NIH/3T3 cells and assessed the level of Sox4 protein expression. The selected PETs were mapped to different regions of the *Sox4* protein-coding (sense) transcript or gene locus. In all experimental groups, expression of *Sox4* transcripts was not significantly different between pcDNA3-empty and mock controls (Fig. 3A). PET2, PET3, PET5 and PET6 were overexpressed in NIH/3T3 cells at 21, 14, 29 and 376 times greater than the mock control, respectively (Fig. 3A). In all the cells that overexpressed PETs, we did not observed any significant changes in *Sox4* sense transcript levels, except for those cells that overexpressed PET2 and PET5, which exhibited about 2.5 and 1.8 times upregulation of the sense transcripts, respectively (Fig. 3A). However, immunoblotting of protein lysates isolated from these cells ($n = 2$) showed no significant changes in Sox4 expression when compared to control, with 0.8-, 1.3-, 1.3-, 0.9- and 0.8-fold changes in PET2-, PET3-, PET5-, PET6- and PcDNA3-transfected cells, respectively (Fig. 3B; see Fig. 3 in [50]). Both RNA and protein analyses suggest that the *Sox4* NATs do not significantly affect the endogenous level of *Sox4* mRNA and protein in the *in vitro* model.

3.4. A novel small RNA, Sox4_sir3, originates from the Sox4 sense transcript

Long dsRNAs can serve as templates for the biogenesis of small RNAs, which function *via* the RNA interference (RNAi) machinery. To answer the question as to whether *Sox4* sense and NATs dsRNA generate functional small RNAs, we generated ~3.7 million small RNA sequences (36 nt) from a mouse E15.5 whole brain using a massively parallel sequencing platform, the Illumina Genome Analyzer II (GSE22653) [53]. Based on the screening of these small RNA sequences, we mapped 7 small RNA sequences with a unique hit to the *Sox4* gene locus where both sense and NATs were expressed (Fig. 4A; see Table 1 in [50]). Interestingly, all 7 small RNA sequences were mapped to the *Sox4* sense transcript and were encountered only once in the dataset suggesting that they could be *Sox4* sense transcript degradation by-products.

To ascertain whether these small RNAs were *Sox4* sense transcript degradation by-products or otherwise, we performed stem loop RT-PCR to evaluate the existence of these small RNAs in the E15.5 mouse brain. The analysis showed only *Sox4*_sir3 small RNA (5′-TCAAGGAC AG CGACAAGATT CCGT-3′; GenBank Accession: HM596744) was specifically amplified, suggesting that this is the only genuine small RNA (Fig. 4B). Next, we searched *Sox4*_sir3 in other high-throughput mouse small RNA sequencing datasets available in the Gene Expression Omnibus (http://www.ncbi.nlm.nih.gov/geo/). We screened 53 high-throughput small RNA sequencing datasets generated from various studies using *Mus musculus* as a model (Series records: GSE20384, GSE19172, GSE17319, GSE7414, GSE5026 and GPL7059). Of these datasets, we found *Sox4*_sir3 was sequenced in only 2 datasets, GSM433295 and GSM475280, which were generated from E18.5 mouse testis small RNAs and Mili-immunoprecipitated adult mouse testis small RNAs, respectively. By considering both laboratory-based and *in silico* analyses of *Sox4*_sir3, we demonstrate that the *Sox4* sense transcript is the origin for the generation of *Sox4*_sir3 novel small

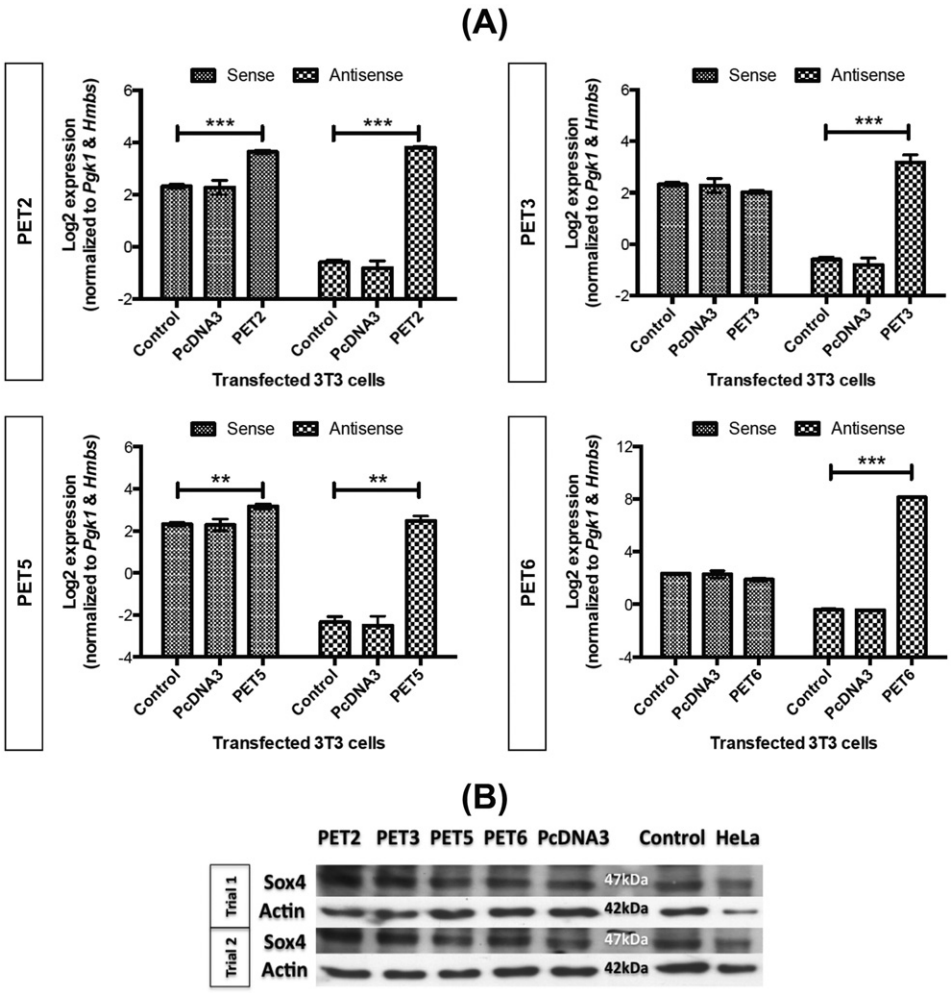

Fig. 3. Overexpression study of PETs 2, 3, 5 and 6 using NIH 3T3 cells. Normalised log2 expression level of *Sox4* sense and NATs in NIH 3T3 cells transfected with reagent only (control), pcDNA3-empty vector (pcDNA3) and individual pcDNA3-PET construct is illustrated in (A) for *Sox4* sense and antisense transcript expression. The western blot analysis based on the antibody against Sox4 or actin protein is shown in (B). An additional lane containing the lysate from HeLa cells was included to serve as positive control. For (A), N = 3 per group and asterisks denote significant level at $^{**}P < 0.01$ and $^{***}P < 0.001$. Error bars denote the standard error of the mean. For (B), N = 2 per group.

RNAs. However, the role of NATs and the mechanism involved in the biogenesis of this small RNA remains unknown.

To determine whether *Sox4*_sir3 small RNA is a novel microRNA (miRNA), we used 150 nt upstream and downstream of *Sox4*_sir3 to look for a potential hairpin stem loop structure using the RNAfold program [57]. Using default thermodynamic prediction parameters [58] and based on the uniform miRNA annotation criteria [59], we did not find any potent hairpin stem loop structures that may have functioned as the *Sox4*_sir3 precursor (Fig. 4C). Therefore, we conclude that *Sox4_sir3* is not a miRNA.

Thus far, we have confirmed that both *Sox4* sense and NATs can form dsRNA in brain cells and *Sox4_sir3* originates from the *Sox4* gene locus, where both sense and NATs are expressed. Since *Sox4*_sir3 is not a miRNA, we speculated that this small RNA could be either an endogenous small interfering RNA (endo-siRNA) or a piRNA that is generated from a long dsRNA *via* a Dicer1-mediated mechanism. To prove our hypothesis, we evaluated the expression of *Sox4*_sir3, *Sox4* sense and NATs in mouse embryonic stem (mES) cells with conditional alleles for *Dicer1* (Fig. 4D). The analysis showed approximately 32-fold down-regulation of *Sox4*_sir3 in Dicer1 null mES cells, compared to cells expressing Dicer1 (Fig. 4D). In addition, we observed no difference in *Sox4* sense transcripts and slight downregulation of NATs expression levels between the two cell types, confirming that *Sox4*_sir3 biogenesis is in fact Dicer1-dependent.

3.5. Sox4 NATs induced the biogenesis of Sox4_sir3 in NIH/3T3 cells

Next, we wanted to determine whether *Sox4* NATs are required for the biogenesis of *Sox4*_sir3. *Sox4*_sir3 is mapped to the coding sequence (CDS) of the *Sox4* sense transcript, and PET6 is the NAT that overlaps this region (see Fig. 4 in [50]). We transfected NIH/3T3 cells with PET6 and assessed the expression level of *Sox4*_sir3. PET3, a *Sox4* NAT that overlaps the 3'UTR of *Sox4* sense transcript, and pcDNA3-empty were transfected as controls. No significant differences between *Sox4* transcripts level between pcDNA3-empty and blank controls were observed in all experimental groups. We observed significant upregulation of both PET3 and PET6 ($P < 0.001$) in the cells and these transcripts did not affect the level of *Sox4* sense transcripts supporting our findings in the previous transfection analyses. Cells overexpressing PET6 showed approximately 63-fold upregulation of *Sox4*_sir3 compared to the mock control (Fig. 5A). However, we did not observe any significant difference in *Sox4*_sir3 expression between cells overexpressing the PET3, and the pcDNA3-empty vector control (Fig. 5A). In addition, *Dicer1*

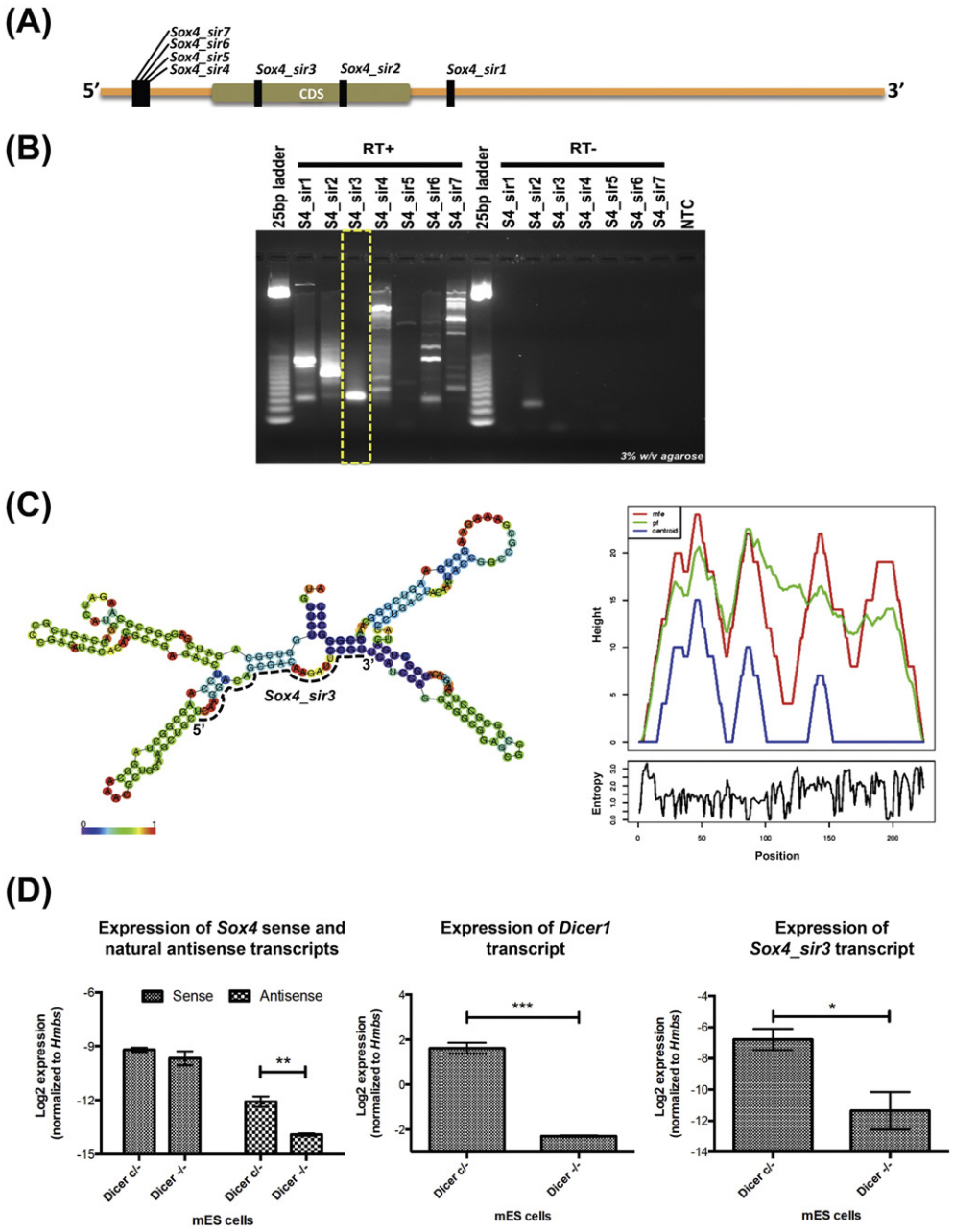

Fig. 4. Identification of *Sox4_sir3* small RNA. (A) Mapping of small RNA sequences that were originated from the *Sox4* gene locus. (B) Validation of small RNAs (S4_sir denotes *Sox4*_sir and NTC denotes no template control) using stemloop RT-PCR method identified *Sox4*_sir3 as the only specific amplicon. (C) RNA fold prediction of sequences 150 nt upstream and downstream of *Sox4_sir3*. The colour in the scale bar presented below to the predicted structure denotes the possibility of base-pairing between nucleotides. The minimum free energy (MFE) structure, the thermodynamic ensemble of RNA structures (pf), the centroid structure (centroid) and the positional entropy for each position are presented in two separated graphs to right of the predicted structure. (D) Normalised log2 expression of *Sox4* sense transcripts, *Sox4* NATs, *Dicer1* and *Sox4_sir3* in mouse embryonic stem (mES) cells with conditional allele for *Dicer1*. Dicer c/— denotes mES cells with Dicer1 activity and Dicer —/— denotes mES cells without Dicer1 activity. For (D), N = 3 per group and asterisks denote significant level at $^{*}P < 0.05$, $^{**}P < 0.01$ and $^{***}P < 0.001$. Error bars denote the standard error of the mean.

expression was not significantly different between transfection groups (Fig. 5B), confirming *Sox4*_sir3 biogenesis is inducible with increased PET6 expression in NIH/3T3 cells (see Fig. 4 in [50]).

3.6. Full-length sequencing of PET6

In general, NATs are expressed in a specific cell type or in certain developmental stages. Full-length sequencing of NATs has been difficult due to their low expression level. To characterise the structure of the PET6 NAT expressed in the brain, we performed full-length sequencing of PET6 NAT isolated from transfected NIH/3T3 cells. We predicted that the cloned PET6 would be expressed and subjected to post-transcriptional processing similar to the naturally occurring PET6. We perform RT-PCR using the primers initially used to amplify PET6 from mouse gDNA. RT-PCR analysis showed amplification of two amplicons, approximately 1.8 kb and 0.6 kb in size (see Supplementary GenBank File and Fig. 4 in [50]). Full-length sequencing confirmed the smaller amplicon as a 0.637 kb spliced transcript variant of PET6. The transcript was spliced from +354 to +1545 (1.187 kb) at the canonical AG…GT acceptor and donor site. A BLASTp homology search and protein domain analysis using Simple Modular Architecture Research Tool (SMART) [60, 61] for open reading frames with greater than 100aa within both

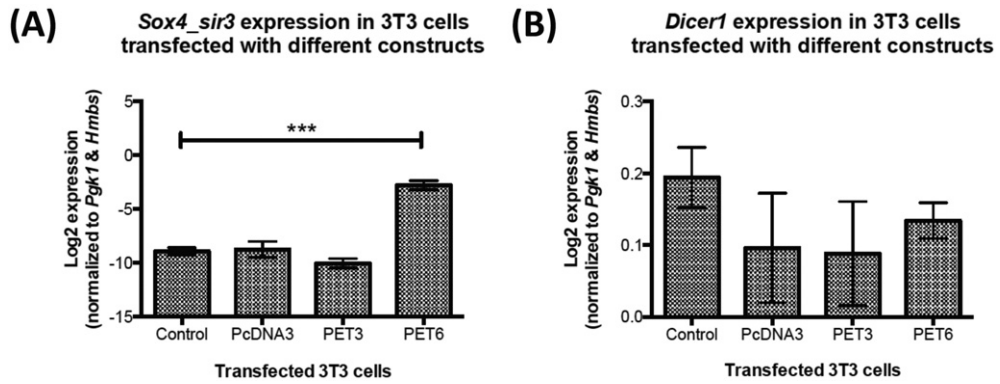

Fig. 5. The effect of PET6 overexpression on *Sox4_sir3* expression. (A) Normalised log2 expression of *Sox4_sir3* small RNA in NIH 3T3 cells transfected with reagent only (mock control), pcDNA3-empty vector (pcDNA3), pcDNA3-PET3 (PET3) and pcDNA3-PET6 (PET6) constructs. (B) Normalised log2 expression of the *Dicer1* transcript in all the NIH 3T3-transcfected cells. N = 3 per group and asterisks dente significant level at $^{***}P < 0.001$. Error bars denote the standard error of the mean.

transcript variants of PET6 did not show any significant protein homologues or functional domains within these transcripts. Irrefutably, these transcripts were not translated into peptides/proteins of known function. These observations led us to ask which transcripts are involved in the generation of *Sox4*_sir3. To address this, we overexpressed both spliced and unspliced PET6 variants in NIH/3T3 cells. We observed upregulation of *Sox4*_sir3 only in cells overexpressing the unspliced variant suggesting the overlapping portion of the transcript is required for *Sox4*_sir3 production (see Figs. 5–6 in [50]). We suspected that the spliced variant of PET6 was an *in vitro* artefact, due to the overexpression of PET6 transcript. In addition, we did not find any RT-qPCR amplification of the spliced variant in various mouse tissues such as brain, heart, spleen, pancreas and skeletal muscle using a pair of intron-spanning primers (data not shown). The unspliced PET6 sequence was termed *Sox4ot1* and was submitted to GenBank with the accession number HM596742.

3.7. Expression pattern of *Sox4_sir3*

To investigate the potential role of *Sox4*_sir3 during brain development, we performed *in situ* hybridisation (ISH) using a commercial Locked-Nucleic Acid (LNA) probe on sections obtained from E11.5, E13.5 and E15.5 whole mouse embryos, and E17.5 and P1.5 whole brains (see Fig. 7 in [50] for LNA-ISH of scramble control). We observed *Sox4*_sir3 expression in the telencephalon and mesencephalon of E11.5 and E13.5 mouse embryos (Fig. 6A-B). The expression was later confined to the ventricular and marginal zones (known as layer I after birth) of the cerebral cortex, ventricular zone, cerebellar anlarge and the granule layer of the olfactory bulb at E15.5 and E17.5 (Fig. 6C–G). At P1.5, *Sox4*_sir3 expression was observed in the diminishing ventricular zone of the cerebral cortex, subventricular zone of the lateral ventricle, layer I of the cerebral cortex, pyramidal layer of the hippocampus, granule cell layer of the dentate gyrus, granule layer of the olfactory

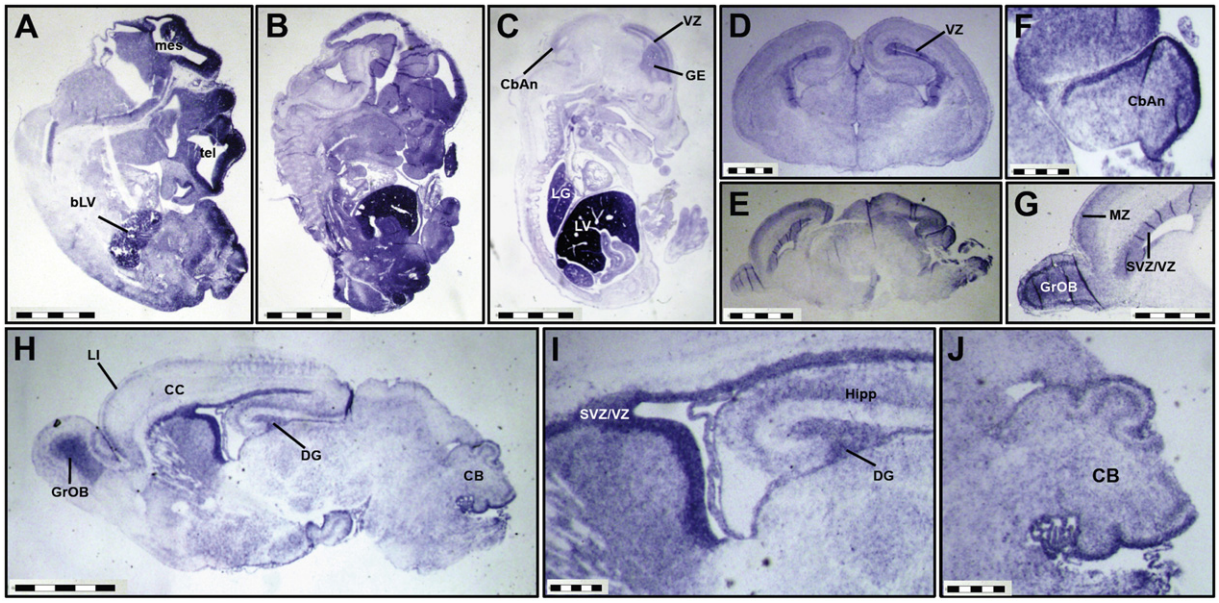

Fig. 6. LNA-ISH of *Sox4_sir3* in whole mouse embryo and brain sections. LNA-ISH of *Sox4_sir3* was performed on whole embryo sections obtained from (A) E11.5, (B) E13.5 and (C) E15.5 embryos. For E17.5 developmental stage, only (D) coronal and (E–G) sagittal whole brain sections were analysed whereas for (H–J) P1.5, sagittal whole brain sections were analysed. 'bLV' = budding liver, 'CB' = cerebellum, 'CbAn' = cerebellar anlage, 'CC' = cerebral cortex, 'DG' = dentate gyrus, 'GE' = ganglionic eminence, 'GrOB' = granule cell layer of the olfactory bulb, 'Hipp' = hippocampus, 'LG' = lungs, 'LI' = layer I of the cerebral cortex, 'LV' = liver, 'mes' = mesencephalon, 'MZ' = marginal zone, 'SVZ' = subventricular zone, 'tel' = telencephalon, 'VZ' = ventricular zone.

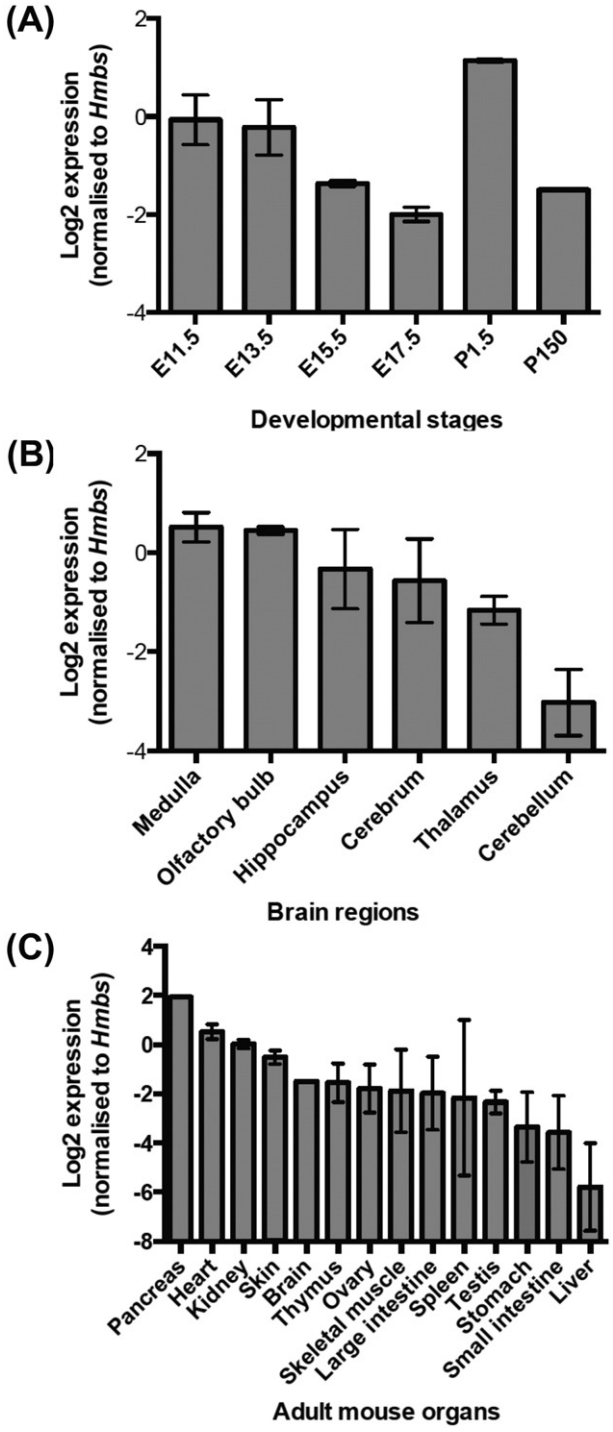

Fig. 7. Stemloop RT-qPCR characterisation of *Sox4_sir3* expression. The characterisation of *Sox4_sir3* expression was performed on (A) whole brains obtained from different developmental stages, (B) different brain regions of adult mice and (C) different adult mouse organs. Data were presented as normalised log2 expression (to *Hmbs* housekeeping gene). For all analyses, N = 2 per sample was used except for the skin, skeletal muscle, spleen and stomach, where N = 3. Error bars denote the standard error of the mean.

bulb and Purkinje cell layers of the cerebellum (Fig. 6H–J). Besides the developing and postnatal brains, *Sox4*_sir3 was strongly expressed in the developing liver between E11.5 and E15.5 (Fig. 6A–C), and the developing lungs at E15.5 (Fig. 6C). These findings show that *Sox4*_sir3 is expressed mainly in germinative zones and specialised neuronal cell layers in the brain as well as the developing lungs and liver suggesting that this small RNA may have a regulatory role in the development or function of these cells or organs, subjecting to further functional validation. Employing multiple mutated LNA probes to serve as negative controls could improve the specificity of the current ISH analysis.

We also performed stemloop RT-qPCR to quantitatively analyse the expression profile of *Sox4*_sir3 in whole brains at different developmental stages, different adult brain regions and different organs of adult mice (Fig. 7; see Supplementary Results in [50] for detail RT-qPCR analysis). We showed *Sox4*_sir3 expression in the whole brain decreases as embryos developed from E11.5 to E17.5. At P1.5, we observed a sudden surge of expression in the whole brain with about a 9-fold increase from E17.5, which then decreased (~6-fold) through to the adult stage at P150 (Fig. 7A). When we analysed different brain regions of adult mice, *Sox4*_sir3 expression was lower in the cerebellum compared to other brain regions (Fig. 7B). When we compared the adult whole brain at P150 with other adult organs, we observed significantly higher *Sox4*_sir3 expression in the heart, kidney and pancreas (Fig. 7C). Interestingly, *Sox4*_sir3 that was found highly expressed in the liver during embryonic development was least expressed among the adult organs screened suggesting that *Sox4*_sir3 may have more important role in embryonic as compared to adult liver development/function.

4. Discussion

In this study, we characterised a cluster of NATs that overlap the *Sox4* sense transcript and described a mechanism related to the ability of these transcripts to form dsRNAs. These dsRNAs served as templates for the production of a novel small RNA, *Sox4*_sir3, *via* a Dicer1-mediated mechanism in the mouse. However, we are unable to classify *Sox4*_sir3 neither as an endogenous small interfering RNA (endo-siRNA) nor a piRNA. Majority of endo-siRNAs are 21–22 nt in size whereas piRNAs are typically 24–32 nt long [40]. *Sox4*_sir3 is 24 nt and this size is not exclusive for either endo-siRNA or piRNA category. However, *Sox4*_sir3 has a 5′ uridine bias, which is a property of piRNAs [62]. The fact that the *Sox4*_sir3 sequence was found to match to those in Mili-immunoprecipitated adult mouse testis small RNAs suggests that this small RNA is very likely to be a piRNA. Although piRNAs have been demonstrated as germline specific small RNAs, recent studies have suggested that they also play a role in nongonadal cells such as mouse hippocampus and neuronal cells in Aplysia [63,64]. Taken into consideration of its size, 5′ biased to uridine and comparative *in silico* analysis results, we suggest that *Sox4*_sir3 is potentially a novel piRNA in the central nervous system. Further investigation based on Ago2- or Mili-immunoprecipitated brain samples is required to confirm the type of small RNA for *Sox4*_sir3.

Biogenesis of small RNAs from dsRNAs is common in viruses and plants where RNA-dependent RNA polymerase (RdRP) plays a pivotal role in catalysing the RNA replication process [65,66]. In mammals, where no RdRP activity has been reported, the mechanism responsible for the generation of dsRNAs remains unclear. In the absence of RdRP, the formation of mammalian dsRNAs have been proposed to ensue *via* other means such as pairing of partially or fully overlapping sense and NATs. The generation of functional small RNAs from dsRNAs in mammals is rare. To date, only limited numbers of small RNAs have been reported as derived from naturally occurring long dsRNAs or retrotransposons found in cultured human cells and murine germ cells during gonadal development [38,67,68]. The present study has demonstrated that *Sox4ot1* induced the overexpression of *Sox4*_sir3, which its functional role is yet to be determined in the mouse or human brain. The direct interaction of *Sox4ot1* and *Sox4* sense transcripts in the *in vitro* system should be evaluated to confirm the proposed biogenesis mechanism of *Sox4*_sir3. *Sox4* sense and NATs were found co-localised in the cytoplasm of brain cells. The observation, however, did not discount the possibility that the transcripts may bind to proteins and protect them from RNAse A digestion. Additional treatment analysis based on RNAse III digestion will add evidence of the presence of a dsRNA complex between *Sox4* sense and NATs transcripts *in situ*.

Unlike other well-characterised NATs such as *Air*, *HOTAIR*, *Evf-2*, *Kcnq1ot1* and *NRON* [12–15,69], *Sox4* NATs were found to form aggregates with its sense counterparts within the cytoplasm, instead of localised within the nucleus. In various incidences, sense-NATs formation in the cytoplasm can inhibit protein translation such as NOS2A in molluscs [70] and FGF-2 in human [71]. In the central nervous system, sense-NATs pairs or NATs alone have been described in the cytoplasm or synaptoneurosomes of neuronal cells with specialised function [63,72,73]. It has been suggested that NATs could exert post-transcriptional regulation on sense transcripts in the brain [74,75]. In another report, cytoplasmic expression of *BACE1* NATs in brain samples from Alzheimer's disease patients has been shown to stabilise the sense transcript by masking the miR-485-5p binding site thus preventing miRNA-induced translational repression [73]. However, the functional role of these NATs in the development or function of neuronal cells has not been conclusively proven. Unlike NATs overlapping *NOS2A*, *FGF-2* and *BACE1*, overexpression of *Sox4* NATs *in vitro* did not significantly affect the Sox4 protein levels but gave rise to *Sox4*_sir3 (*via* PET6), which its origin was mapped to sense transcript. The low copy number of *Sox4_sir3* in our sequencing dataset is not an uncommon phenomenon. A total of 23 known miRNAs were also found to have a single read count in the dataset [53], suggesting that low copy number sequences including *Sox4_sir3* may not necessarily non-functional. Overexpression analyses performed on the NIH/3T3 cell line in the present study were limited to the evaluation of endogenous *Sox4* transcript or protein expression levels within the cells. Additional functional characterisation of NATs *in vivo* is required to conclude the effect of *Sox4* NATs especially the *Sox4ot1* on *Sox4* mRNA/protein expression within the brain.

Expression of *Sox4*_sir3 is spatiotemporally regulated during embryogenesis between E11.5 and E17.5. In E11.5 embryos, *Sox4*_sir3 expression was specific to the liver bud and the telencephalon. Between E11.5 and E15.5, the developing liver bud is a major site for foetal haematopoiesis. During these stages of development, haematopoietic stem cells (HSCs) increased exponentially in the foetal liver [76,77] followed by mobilisation of HSCs to spleen and bone marrow after E16 [77]. Although *Sox4*_sir3 expression was very specific to the liver, its role in liver development or foetal haematopoiesis remains unclear and requires further experimental validation.

Temporal analysis of the whole mouse brain showed a sudden surge of *Sox4*_sir3 expression at the P1.5 developmental stage. In rodent, gliogenesis is at peak right afterbirth [78]. Increased expression of *Sox4*_sir3 at P1.5 mainly in the subventricular/ventricular zones may be associated with active gliogenesis activities at the region. The *Sox4_sir3* expression was also found specifically in the marginal zone and layer I of the embryonic and postnatal mouse cerebral cortices, respectively. Interestingly, these regions are predominated by Cajal–Retzius neurones that secrete Reelin, a protein involved in establishing early neuronal circuitry, cortical lamination and cortical evolution [79,80]. Therefore, the role of *Sox4*_sir3 in the development and function of Cajal–Retzius neurones should be further evaluated.

We have demonstrated that the overexpression of selected *Sox4* NATs did not exert significant changes to *Sox4* transcripts or protein levels in NIH/3T3 cell line but a novel *Sox4* NAT, *Sox4ot1*, may induce the production of a piRNA-like small RNA, *Sox4*_sir3 in the mammalian cell system. Based on our observations in RNA FISH and LNA ISH experiments, we propose that the mechanism of *Sox4_sir3* biogenesis in the NIH/3T3 artificial model involves the formation of dsRNA between *Sox4* sense and *Sox4ot1* followed by Dicer1 dependent production of *Sox4_sir3*. Our study, however, would be better enhanced with further functional characterisation of *Sox4ot1* and *Sox4_sir3* using primary neuronal or glial cells to deduce the role of both NAT and small RNA throughout neurodevelopment. In addition, *Sox4_sir3* may involve in mechanisms that are not associated with Sox4 protein function and therefore the role of the small RNA should be investigated from different perspectives. Our preliminary findings, when further proven or characterised *in vivo*, will have significant implications on our understanding of NATs role in small RNAs biogenesis. About 20% of well-defined proteins have at least one overlapping transcript [81], thus interactions of these sense-antisense RNA pairs may contribute to unorthodox production of small RNAs that function as secondary regulatory RNAs, hence explaining the missing link and the wide-spectrum of transcript-protein relationship in various complex processes involved in organismal development.

Acknowledgement

This work was supported by National Health and Medical Research Council fellowships [171601 and 461204 to H.S.S.], National Health and Medical Research Council Grants [219176, 257501 and 257529 to H.S.S.], Malaysian Ministry of Science, Technology & Innovation (MOSTI) Science Fund (02-01-04-SF2336 to K.H.L.), Malaysian Ministry of Higher Education Fundamental Research Grant Scheme (04-01-15-1663FR to P.S.C.), and a fellowship from Pfizer Australia [to P.Q.T.]. We would like to express our gratitude to Professor Greg Goodall from SA Pathology at IMVS, South Australia for his constructive comments and views regarding the manuscript.

References

[1] E.S. Lander, L.M. Linton, B. Birren, C. Nusbaum, M.C. Zody, J. Baldwin, et al., Initial sequencing and analysis of the human genome, Nature 409 (2001) 860–921, http://dx.doi.org/10.1038/35057062.
[2] J.C. Venter, M.D. Adams, E.W. Myers, P.W. Li, R.J. Mural, G.G. Sutton, et al., The sequence of the human genome, Science 291 (2001) 1304–1351, http://dx.doi.org/10.1126/science.1058040.
[3] ENCODE Project Consortium, E. Birney, J.A. Stamatoyannopoulos, A. Dutta, R. Guigó, T.R. Gingeras, et al., Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project, Nature 447 (2007) 799–816, http://dx.doi.org/10.1038/nature05874.
[4] P. Bertone, V. Stolc, T.E. Royce, J.S. Rozowsky, A.E. Urban, X. Zhu, et al., Global identification of human transcribed sequences with genome tiling arrays, Science 306 (2004) 2242–2246, http://dx.doi.org/10.1126/science.1103388.
[5] P. Carninci, T. Kasukawa, S. Katayama, J. Gough, M.C. Frith, N. Maeda, et al., The transcriptional landscape of the mammalian genome, Science 309 (2005) 1559–1563, http://dx.doi.org/10.1126/science.1112014.
[6] J. Cheng, P. Kapranov, J. Drenkow, S. Dike, S. Brubaker, S. Patel, et al., Transcriptional maps of 10 human chromosomes at 5-nucleotide resolution, Science 308 (2005) 1149–1154, http://dx.doi.org/10.1126/science.1108625.
[7] N.D. Trinklein, U. Karaöz, J. Wu, A. Halees, S. Force Aldred, P.J. Collins, et al., Integrated analysis of experimental data sets reveals many novel promoters in 1% of the human genome, Genome Res. 17 (2007) 720–731, http://dx.doi.org/10.1101/NA.
[8] M.E. Fahey, T.F. Moore, D.G. Higgins, Overlapping antisense transcription in the human genome, Comp. Funct. Genomics 3 (2002) 244–253, http://dx.doi.org/10.1002/cfg.173.
[9] J. Shendure, G.M. Church, Computational discovery of sense-antisense transcription in the human and mouse genomes, Genome Biol. 3 (2002), RESEARCH0044.
[10] H. Kiyosawa, I. Yamanaka, N. Osato, S. Kondo, Y. Hayashizaki, RIKEN GER Group, et al., Antisense transcripts with FANTOM2 clone set and their implications for gene regulation, Genome Res. 13 (2003) 1324–1334, http://dx.doi.org/10.1101/gr.982903.
[11] R. Yelin, D. Dahary, R. Sorek, E.Y. Levanon, O. Goldstein, A. Shoshan, et al., Widespread occurrence of antisense transcription in the human genome, Nat. Biotechnol. 21 (2003) 379–386, http://dx.doi.org/10.1038/nbt808.
[12] J. Feng, C. Bi, B.S. Clark, R. Mady, P. Shah, J.D. Kohtz, The Evf-2 noncoding RNA is transcribed from the Dlx-5/6 ultraconserved region and functions as a Dlx-2 transcriptional coactivator, Genes Dev. 20 (2006) 1470–1484, http://dx.doi.org/10.1101/gad.1416106.
[13] F. Sleutels, R. Zwart, D.P. Barlow, The non-coding Air RNA is required for silencing autosomal imprinted genes, Nature 415 (2002) 810–813, http://dx.doi.org/10.1038/415810a.
[14] J.L. Rinn, M. Kertesz, J.K. Wang, S.L. Squazzo, X. Xu, S.A. Brugmann, et al., Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs, Cell 129 (2007) 1311–1323, http://dx.doi.org/10.1016/j.cell.2007.05.022.
[15] R.R. Pandey, T. Mondal, F. Mohammad, S. Enroth, L. Redrup, J. Komorowski, et al., Kcnq1ot1 antisense noncoding RNA mediates lineage-specific transcriptional silencing through chromatin-level regulation, Mol. Cell 32 (2008) 232–246, http://dx.doi.org/10.1016/j.molcel.2008.08.022.
[16] P.D. Ladd, L.E. Smith, N.A. Rabaia, J.M. Moore, S.A. Georges, R.S. Hansen, et al., An antisense transcript spanning the CGG repeat region of FMR1 is upregulated in premutation carriers but silenced in full mutation individuals, Hum. Mol. Genet. 16 (2007) 3174–3187, http://dx.doi.org/10.1093/hmg/ddm293.
[17] C.A. Thrash-Bingham, K.D. Tartof, aHIF: a natural antisense transcript overexpressed in human renal cancer and during hypoxia, J. Natl. Cancer Inst. 91 (1999) 143–151.

[18] T. Yamamoto, Y. Manome, M. Nakamura, N. Tanigawa, Downregulation of survivin expression by induction of the effector cell protease receptor-1 reduces tumor growth potential and results in an increased sensitivity to anticancer agents in human colon cancer, Eur. J. Cancer 38 (2002) 2316–2324.
[19] N. Berteaux, N. Aptel, G. Cathala, C. Genton, J. Coll, A. Daccache, et al., A novel H19 antisense RNA overexpressed in breast cancer contributes to paternal IGF2 expression, Mol. Cell. Biol. 28 (2008) 6731–6745, http://dx.doi.org/10.1128/MCB.02103-07.
[20] S. Capaccioli, A. Quattrone, N. Schiavone, A. Calastretti, E. Copreni, A. Bevilacqua, et al., A bcl-2/IgH antisense transcript deregulates bcl-2 gene expression in human follicular lymphoma t(14;18) cell lines, Oncogene 13 (1996) 105–115.
[21] N.J. Smilinich, C.D. Day, G.V. Fitzpatrick, G.M. Caldwell, A.C. Lossie, P.R. Cooper, et al., A maternally methylated CpG island in KvLQT1 is associated with an antisense paternal transcript and loss of imprinting in Beckwith–Wiedemann syndrome, Proc. Natl. Acad. Sci. U. S. A. 96 (1999) 8064–8069.
[22] C. Tufarelli, J.A.S. Stanley, D. Garrick, J.A. Sharpe, H. Ayyub, W.G. Wood, et al., Transcription of antisense RNA leading to gene silencing and methylation as a novel cause of human genetic disease, Nat. Genet. 34 (2003) 157–165, http://dx.doi.org/10.1038/ng1157.
[23] K.-H. Ling, C.A. Hewitt, T. Beissbarth, L. Hyde, K. Banerjee, P.-S. Cheah, et al., Molecular networks involved in mouse cerebral corticogenesis and spatio-temporal regulation of Sox4 and Sox11 novel antisense transcripts revealed by transcriptome profiling, Genome Biol. 10 (2009) R104, http://dx.doi.org/10.1186/gb-2009-10-10-r104.
[24] J.C. Kiefer, Back to basics: Sox genes, Dev. Dyn. 236 (2007) 2356–2366, http://dx.doi.org/10.1002/dvdy.21218.
[25] V. Lefebvre, B. Dumitriu, A. Penzo-Méndez, Y. Han, B. Pallavi, Control of cell fate and differentiation by Sry-related high-mobility-group box (Sox) transcription factors, Int. J. Biochem. Cell Biol. 39 (2007) 2195–2214, http://dx.doi.org/10.1016/j.biocel.2007.05.019.
[26] Y. Kamachi, M. Uchikawa, H. Kondoh, Pairing SOX off: with partners in the regulation of embryonic development, Trends Genet. 16 (2000) 182–187.
[27] M. van de Wetering, M. Oosterwegel, K. van Norren, H. Clevers, Sox-4, an Sry-like HMG box protein, is a transcriptional activator in lymphocytes, EMBO J. 12 (1993) 3847–3854.
[28] M.W. Schilham, M.A. Oosterwegel, P. Moerer, J. Ya, P.A. de Boer, M. van de Wetering, et al., Defects in cardiac outflow tract formation and pro-B-lymphocyte expansion in mice lacking Sox-4, Nature 380 (1996) 711–714, http://dx.doi.org/10.1038/380711a0.
[29] M. Cheung, M. Abu-Elmagd, H. Clevers, P.J. Scotting, Roles of Sox4 in central nervous system development, Brain Res. Mol. Brain Res. 79 (2000) 180–191.
[30] M. Bergsland, M. Werme, M. Malewicz, T. Perlmann, J. Muhr, The establishment of neuronal properties is controlled by Sox4 and Sox11, Genes Dev. 20 (2006) 3475–3486, http://dx.doi.org/10.1101/gad.403406.
[31] M.R. Potzner, C. Griffel, E. Lütjen-Drecoll, M.R. Bösl, M. Wegner, E. Sock, Prolonged Sox4 expression in oligodendrocytes interferes with normal myelination in the central nervous system, Mol. Cell. Biol. 27 (2007) 5316–5326, http://dx.doi.org/10.1128/MCB.00339-07.
[32] C.S. Bangur, A. Switzer, L. Fan, M.J. Marton, M.R. Meyer, T. Wang, Identification of genes over-expressed in small cell lung carcinoma using suppression subtractive hybridization and cDNA microarray expression analysis, Oncogene 21 (2002) 3814–3825, http://dx.doi.org/10.1038/sj.onc.1205480.
[33] C.-J. Lee, V.J. Appleby, A.T. Orme, W.-I. Chan, P.J. Scotting, Differential expression of SOX4 and SOX11 in medulloblastoma, J. Neuro-Oncol. 57 (2002) 201–214.
[34] M. Aaboe, K. Birkenkamp-Demtroder, C. Wiuf, F.B. Sørensen, T. Thykjaer, G. Sauter, et al., SOX4 expression in bladder carcinoma: clinical aspects and in vitro functional characterization, Cancer Res. 66 (2006) 3434–3442, http://dx.doi.org/10.1158/0008-5472.CAN-05-3456.
[35] K.E. Boyd, Y.-Y. Xiao, K. Fan, A. Poholek, N.G. Copeland, N.A. Jenkins, et al., Sox4 cooperates with Evi1 in AKXD-23 myeloid tumors via transactivation of proviral LTR, Blood 107 (2006) 733–741, http://dx.doi.org/10.1182/blood-2003-05-1626.
[36] P. Liu, S. Ramachandran, M. Ali Seyed, C.D. Scharer, N. Laycock, W.B. Dalton, et al., Sex-determining region Y box 4 is a transforming oncogene in human prostate cancer cells, Cancer Res. 66 (2006) 4011–4019, http://dx.doi.org/10.1158/0008-5472.CAN-05-3055.
[37] D. Sinner, J.J. Kordich, J.R. Spence, R. Opoka, S. Rankin, S.-C.J. Lin, et al., Sox17 and Sox4 differentially regulate beta-catenin/T-cell factor activity and proliferation of colon carcinoma cells, Mol. Cell. Biol. 27 (2007) 7802–7815, http://dx.doi.org/10.1128/MCB.02179-06.
[38] T. Watanabe, Y. Totoki, A. Toyoda, M. Kaneda, S. Kuramochi-Miyagawa, Y. Obata, et al., Endogenous siRNAs from naturally formed dsRNAs regulate transcripts in mouse oocytes, Nature 453 (2008) 539–543, http://dx.doi.org/10.1038/nature06908.
[39] C. Klattenhoff, W. Theurkauf, Biogenesis and germline functions of piRNAs, Development 135 (2008) 3–9, http://dx.doi.org/10.1242/dev.006486.
[40] T. Thomson, H. Lin, The biogenesis and function of PIWI proteins and piRNAs: progress and prospect, Annu. Rev. Cell Dev. Biol. 25 (2009) 355–376, http://dx.doi.org/10.1146/annurev.cellbio.24.110707.175327.
[41] D.P. Bartel, MicroRNAs: genomics, biogenesis, mechanism, and function, Cell 116 (2004) 281–297.
[42] A. Aravin, D. Gaidatzis, S. Pfeffer, M. Lagos-Quintana, P. Landgraf, N. Iovino, et al., A novel class of small RNAs bind to MILI protein in mouse testes, Nature 442 (2006) 203–207, http://dx.doi.org/10.1038/nature04916.
[43] A.A. Aravin, R. Sachidanandam, A. Girard, K. Fejes-Toth, G.J. Hannon, Developmentally regulated piRNA clusters implicate MILI in transposon control, Science 316 (2007) 744–747, http://dx.doi.org/10.1126/science.1142612.
[44] J. Brennecke, A. Stark, R.B. Russell, S.M. Cohen, Principles of microRNA-target recognition, PLoS Biol. 3 (2005), e85http://dx.doi.org/10.1371/journal.pbio.0030085.
[45] D.P. Bartel, MicroRNAs: target recognition and regulatory functions, Cell 136 (2009) 215–233, http://dx.doi.org/10.1016/j.cell.2009.01.002.
[46] S.A. Shabalina, E.V. Koonin, Origins and evolution of eukaryotic RNA interference, Trends Ecol. Evol. (Amst.) 23 (2008) 578–587, http://dx.doi.org/10.1016/j.tree.2008.06.005.
[47] H.R. Chiang, L.W. Schoenfeld, J.G. Ruby, V.C. Auyeung, N. Spies, D. Baek, et al., Mammalian microRNAs: experimental evaluation of novel and previously annotated genes, Genes Dev. 24 (2010) 992–1009, http://dx.doi.org/10.1101/gad.1884710.
[48] N.C.T. Schopman, S. Heynen, J. Haasnoot, B. Berkhout, A miRNA-tRNA mix-up: tRNA origin of proposed miRNA, RNA Biol. 7 (2010) 573–576, http://dx.doi.org/10.4161/rna.7.4.13141.
[49] L. Soreq, N. Salomonis, M. Bronstein, D.S. Greenberg, Z. Israel, H. Bergman, et al., Small RNA sequencing-microarray analyses in Parkinson leukocytes reveal deep brain stimulation-induced splicing changes that classify brain region transcriptomes, Front. Mol. Neurosci. 6 (2013) 10, http://dx.doi.org/10.3389/fnmol.2013.00010.
[50] K.H. Ling, P.J. Brautigan, S. Moore, R. Fraser, M.P.Y. Leong, J.W. Leong, et al., In depth analysis of the Sox4 gene locus that consists of sense and antisense transcripts, Data in Brief (2016).
[51] K. Rutherford, J. Parkhill, J. Crook, T. Horsnell, P. Rice, M.A. Rajandream, et al., Artemis: sequence visualization and annotation, Bioinformatics 16 (2000) 944–945.
[52] K.H. Ling, C.A. Hewitt, T. Beissbarth, L. Hyde, P.-S. Cheah, G.K. Smyth, et al., Spatio-temporal regulation of multiple overlapping sense and novel natural antisense transcripts at the Nrgn and Camk2n1 gene loci during mouse cerebral corticogenesis, Cereb. Cortex 21 (2011) 683–697, http://dx.doi.org/10.1093/cercor/bhq141.
[53] K.H. Ling, P.J. Brautigan, C.N. Hahn, T. Daish, J.R. Rayner, P.-S. Cheah, et al., Deep sequencing analysis of the developing mouse brain reveals a novel microRNA, BMC Genomics 12 (2011) 176, http://dx.doi.org/10.1186/1471-2164-12-176.
[55] D.M. Mattiske, L. Han, J.R. Mann, Meiotic maturation failure induced by DICER1 deficiency is derived from primary oocyte ooplasm, Reproduction 137 (2009) 625–632, http://dx.doi.org/10.1530/REP-08-0475.
[56] E. Beaudoing, S. Freier, J.R. Wyatt, J.M. Claverie, D. Gautheret, Patterns of variant polyadenylation signal usage in human genes, Genome Res. 10 (2000) 1001–1010.
[57] A.R. Gruber, R. Lorenz, S.H. Bernhart, R. Neuböck, I.L. Hofacker, The Vienna RNA website, Nucleic Acids Res. 36 (2008) W70–W74, http://dx.doi.org/10.1093/nar/gkn188.
[58] D.H. Mathews, J. Sabina, M. Zuker, D.H. Turner, Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure, J. Mol. Biol. 288 (1999) 911–940, http://dx.doi.org/10.1006/jmbi.1999.2700.
[59] V. Ambros, B. Bartel, D.P. Bartel, C.B. Burge, J.C. Carrington, X. Chen, et al., A uniform system for microRNA annotation, RNA 9 (2003) 277–279, http://dx.doi.org/10.1261/rna.2183803.
[60] I. Letunic, T. Doerks, P. Bork, SMART 6: recent updates and new developments, Nucleic Acids Res. 37 (2009) D229–D232, http://dx.doi.org/10.1093/nar/gkn808.
[61] I. Letunic, T. Doerks, P. Bork, SMART: recent updates, new developments and status in 2015, Nucleic Acids Res. 43 (2015) D257–D260, http://dx.doi.org/10.1093/nar/gku949.
[62] J. Brennecke, A.A. Aravin, A. Stark, M. Dus, M. Kellis, R. Sachidanandam, et al., Discrete small RNA-generating loci as master regulators of transposon activity in Drosophila, Cell 128 (2007) 1089–1103, http://dx.doi.org/10.1016/j.cell.2007.01.043.
[63] N.R. Smalheiser, G. Lugli, V.I. Torvik, N. Mise, R. Ikeda, K. Abe, Natural antisense transcripts are co-expressed with sense mRNAs in synaptoneurosomes of adult mouse forebrain, Neurosci. Res. 62 (2008) 236–239, http://dx.doi.org/10.1016/j.neures.2008.08.010.
[64] P. Rajasethupathy, I. Antonov, R. Sheridan, S. Frey, C. Sander, T. Tuschl, et al., A role for neuronal piRNAs in the epigenetic control of memory-related synaptic plasticity, Cell 149 (2012) 693–707, http://dx.doi.org/10.1016/j.cell.2012.02.057.
[65] P. Ahlquist, RNA-dependent RNA polymerases, viruses, and RNA silencing, Science 296 (2002) 1270–1273, http://dx.doi.org/10.1126/science.1069132.
[66] T. Sugiyama, H. Cam, A. Verdel, D. Moazed, S.I.S. Grewal, RNA-dependent RNA polymerase is an essential component of a self-enforcing loop coupling heterochromatin assembly to siRNA production, Proc. Natl. Acad. Sci. U. S. A. 102 (2005) 152–157, http://dx.doi.org/10.1073/pnas.0407641102.
[67] T. Watanabe, A. Takeda, T. Tsukiyama, K. Mise, T. Okuno, H. Sasaki, et al., Identification and characterization of two novel classes of small RNAs in the mouse germline: retrotransposon-derived siRNAs in oocytes and germline small RNAs in testes, Genes Dev. 20 (2006) 1732–1743, http://dx.doi.org/10.1101/gad.1425706.
[68] N. Yang, H.H. Kazazian, L1 retrotransposition is suppressed by endogenously encoded small interfering RNAs in human cultured cells, Nat. Struct. Mol. Biol. 13 (2006) 763–771, http://dx.doi.org/10.1038/nsmb1141.
[69] A.T. Willingham, A.P. Orth, S. Batalov, E.C. Peters, B.G. Wen, P. Aza-Blanc, et al., A strategy for probing the function of noncoding RNAs finds a repressor of NFAT, Science 309 (2005) 1570–1573, http://dx.doi.org/10.1126/science.1115901.
[70] S.A. Korneev, J.H. Park, M. O'Shea, Neuronal expression of neural nitric oxide synthase (nNOS) protein is suppressed by an antisense RNA transcribed from an NOS pseudogene, J. Neurosci. 19 (1999) 7711–7720.
[71] A.W. Li, P.R. Murphy, Expression of alternatively spliced FGF-2 antisense RNA transcripts in the central nervous system: regulation of FGF-2 mRNA translation, Mol. Cell. Endocrinol. 162 (2000) 69–78.
[72] M. Chiba, M. Kubo, T. Miura, T. Sato, A.H. Rezaeian, H. Kiyosawa, et al., Localization of sense and antisense transcripts of Prdx2 gene in mouse tissues, Cytogenet. Genome Res. 121 (2008) 222–231, http://dx.doi.org/10.1159/000138889.
[73] M.A. Faghihi, M. Zhang, J. Huang, F. Modarresi, M.P. Van der Brug, M.A. Nalls, et al., Evidence for natural antisense transcript-mediated inhibition of microRNA function, Genome Biol. 11 (2010) R56, http://dx.doi.org/10.1186/gb-2010-11-5-r56.
[74] R. Parenti, S. Paratore, A. Torrisi, S. Cavallaro, A natural antisense transcript against Rad18, specifically expressed in neurons and upregulated during beta-amyloid-

induced apoptosis, Eur. J. Neurosci. 26 (2007) 2444–2457, http://dx.doi.org/10.1111/j.1460-9568.2007.05864.x.
[75] G. Spigoni, C. Gedressi, A. Mallamaci, Regulation of Emx2 expression by antisense transcripts in murine cortico-cerebral precursors, PLoS ONE 5 (2010), e8658http://dx.doi.org/10.1371/journal.pone.0008658.
[76] K. Ikuta, I.L. Weissman, Evidence that hematopoietic stem cells express mouse c-kit but do not depend on steel factor for their generation, Proc. Natl. Acad. Sci. U. S. A. 89 (1992) 1502–1506.
[77] S.J. Morrison, H.D. Hemmati, A.M. Wandycz, I.L. Weissman, The purification and characterization of fetal liver hematopoietic stem cells, Proc. Natl. Acad. Sci. U. S. A. 92 (1995) 10302–10306.
[78] W.-P. Ge, A. Miyawaki, F.H. Gage, Y.N. Jan, L.Y. Jan, Local generation of glia is a major astrocyte source in postnatal cortex, Nature 484 (2012) 376–380, http://dx.doi.org/10.1038/nature10959.
[79] A. Aguiló, T.H. Schwartz, V.S. Kumar, Z.A. Peterlin, A. Tsiola, E. Soriano, et al., Involvement of cajal–retzius neurons in spontaneous correlated activity of embryonic and postnatal layer 1 from wild-type and reeler mice, J. Neurosci. 19 (1999) 10856–10868.
[80] I. Bar, C. Lambert de Rouvroit, A.M. Goffinet, The evolution of cortical development. An hypothesis based on the role of the Reelin signaling pathway, Trends Neurosci. 23 (2000) 633–638.
[81] J. Chen, M. Sun, W.J. Kent, X. Huang, H. Xie, W. Wang, et al., Over 20% of human transcripts might form sense-antisense pairs, Nucleic Acids Res. 32 (2004) 4812–4820, http://dx.doi.org/10.1093/nar/gkh818.
[82] J. Ip, P. Canham, K.H. Choo, Y. Inaba, S.A. Jacobs, P. Kalitsis, D.M. Mattiske, J. Ng, R. Saffery, N.C. Wong, L.H. Wong, J.R. Mann, Normal DNA methylation dynamics in DICER1-deficient mouse embryonic stem cells, PLoS Genet. 8 (9) (2012), e1002919, http://dx.doi.org/10.1371/journal.pgen.1002919.