2274–2282 Nucleic Acids Research, 2009, Vol. 37, No. 7 Published online 23 February 2009
doi:10.1093/nar/gkp088

Strand selective generation of endo-siRNAs from the Na/phosphate transporter gene *Slc34a1* in murine tissues

Mark Carlile[1], Daniel Swan[2], Kelly Jackson[1], Keziah Preston-Fayers[1], Benoit Ballester[3], Paul Flicek[3] and Andreas Werner[1,*]

[1]RNA Biology Group and Epithelial Research Group, [2]Bioinformatics Support Unit, Institute for Cell and Molecular Biosciences, Newcastle University, Framlington Place, Newcastle NE2 4HH and [3]European Bioinformatics Institute, Wellcome Trust Genome Campus, Hinxton, CB10 1SD, UK

Received November 3, 2008; Revised January 23, 2009; Accepted February 3, 2009

ABSTRACT

Natural antisense transcripts (NATs) are important regulators of gene expression. Recently, a link between antisense transcription and the formation of endo-siRNAs has emerged. We investigated the bi-directionally transcribed Na/phosphate cotransporter gene (*Slc34a1*) under the aspect of endo-siRNA processing. Mouse *Slc34a1* produces an antisense transcript that represents an alternative splice product of the *Pfn3* gene located downstream of *Slc34a1*. The antisense transcript is prominently found in testis and in kidney. Co-expression of *in vitro* synthesized sense/antisense transcripts in *Xenopus* oocytes indicated processing of the overlapping transcripts into endo-siRNAs in the nucleus. Truncation experiments revealed that an overlap of at least 29 base-pairs is required to induce processing. We detected endo-siRNAs in mouse tissues that co express *Slc34a1* sense/antisense transcripts by northern blotting. The orientation of endo-siRNAs was tissue specific in mouse kidney and testis. In kidney where the Na/phosphate cotransporter fulfils its physiological function endo-siRNAs complementary to the NAT were detected, in testis both orientations were found. Considering the wide spread expression of NATs and the gene silencing potential of endo-siRNAs we hypothesized a genome-wide link between antisense transcription and monoallelic expression. Significant correlation between random imprinting and antisense transcription could indeed be established. Our findings suggest a novel, more general role for NATs in gene regulation.

INTRODUCTION

Natural antisense transcripts (NATs) represent a widespread phenomenon observed in all organisms (1). Sense-antisense transcript pairs comprise conventionally a protein coding sense mRNA, generally higher and more widely expressed and better characterized than the corresponding NAT, and a regulatory, often non coding antisense transcript (2). In prokaryotes NATs control plasmid copy numbers whereas NATs are involved in viral defense and stress responses in plants (3,4). In higher vertebrates, especially mouse and human, the biological role of NATs is controversial (5–7). There is clear evidence that NATs play an essential role in the epigenetic silencing of mono allelically expressed gene clusters such as parentally imprinted genes, immunoglobulin genes or odorant receptor genes (8,9). However, the total number of NATs exceeds those with an established function by more than an order of magnitude (10). This raises the question whether the myriad of uncharacterized NATs represent specific regulators for the related individual genes or whether a general regulatory concept for NATs remains to be established (11). There is increasing evidence that indeed both scenarios apply. The impact of NATs on the physiological regulation of the corresponding sense transcript and the encoded protein has been demonstrated for Msx1, β secretase and both thyroid and erythropoietin receptors to name just a few examples (12–15). This approach to gene regulation by NATs has been comprehensively reviewed by Beiter *et al.* (1). Recently, a more general role of NATs in quality control of transcripts has been suggested (11).

Initial efforts to investigate the expression of NATs on a genome wide basis selected for spliced and/or polyadenylated transcripts in order to exclude wrongly oriented ESTs (16,17). The total number of antisense transcripts varied from study to study due to experimental differences; however, they all agreed on the fact that

*To whom correspondence should be addressed. Tel: +44 191 222 6990; Fax: +44 191 222 7424; Email: andreas.werner@ncl.ac.uk

sense-antisense transcript pairs with exonic complementarity are under represented on the X-chromosome as compared to autosomes. If transcript pairs with exonic complementarity were excluded from the analysis, the remaining sense transcript/NATs pairs were found to be distributed evenly throughout the genome (16,18). These hallmarks suggest that a significant group of NATs experience similar evolutionary constraints. The regulatory process related to this special group of NATs most likely relies on complementary RNA sequences (7,11).

Recently, a link between NATs expression and the formation of endogenous short interfering RNAs (endo-siRNAs) has been established. Both large scale sequencing and investigations into specific bi-directionally transcribed genes identified endo-siRNAs that matched the complementary regions of sense transcript/NAT pairs (19–21). In all cases, the orientation of the single-stranded endo-siRNAs was non-random. Watanabe *et al*. detected sense oriented endo-siRNAs (complementary to the NAT) in mouse oocytes (21). Carlile *et al*. reported a switch from antisense oriented to sense oriented siRNAs in zebrafish embryonic development between 48 and 72 h post fertilization (19).

We took an integrative approach with the aim of extrapolating the findings from a single bi-directionally transcribed gene, *Slc34a1*|*Pfn3*, to a more general concept of gene regulation by NATs. Importantly, we could confirm the production of NATs-related endo-siRNAs and the non-random selection of endo-siRNA strands. Based on these findings we hypothesized and confirmed a general link between NATs and random monoallelic expression.

MATERIALS AND METHODS

Animals

BALB/c mice were bred in-house. Young male and female animals were killed by cervical dislocation according to the home office schedule 1 procedure. Tissues of interest were quickly removed, rinsed in ice cold PBS and frozen in liquid nitrogen. Frogs were obtained from the African *Xenopus* Facility (South Africa). *Xenopus* were anaesthetized by immersion into ice-cold tricaine solution (Sigma) and killed by decapitation before the removal of the oocytes according to registered procedures. The oocytes were surgically removed, rinsed in ORII solution (NaCl 82.5 mM, KCl 2 mM, $MgCl_2$ 1 mM, HEPES pH 7.4, 5 mM) followed by collagenase treatment (2 mg/ml in ORII) to remove the follicular cell layer. Oocytes were stored in modified Barth's solution (NaCl 88 mM, KCl 1 mM, $MgSO_4$ 0.82 mM, $CaCl_2$ 0.41 mM, $Ca(NO_3)_2$ 0.33 mM, HEPES/Tris pH 7.5, 10 mM, $NaHCO_3$ 2.4 mM) at 18°C. The care and use of all experimental animals was carried out in accordance with the guidelines of the UK Homeoffice.

Expression in Xenopus oocytes

cRNA was produced from *Hind*III or *EcoR*V linearized plasmids using the mMESSAGE mMACHINE kit (Ambion). Alternatively, RNA was synthesized directly from amplified DNA fragments that contained a T_7-binding site added to one of the primers used for the PCR. The RNA was purified by LiCl precipitation and concentration was determined spectrophotometrically. Routinely, between 2 and 10 ng of cRNA in 50 nl of water were injected into oocytes. For nuclear injections, the injection volume was decreased to 10 nl without changing the amount of RNA. The oocytes were kept at 18°C in modified Barth's solution up to 3 days before the assays. All experiments were done with oocytes from at least three different frogs with comparable results.

Isolation of nucleic acids

Mouse tissues (<100 mg) were ground to powder in liquid nitrogen and added to 1 ml of Tri-reagent (Sigma). Homogenization was completed by passing the slurry repeatedly through 21 and 25 G injection needles. RNA extraction was then performed according to the supplier's protocol. The integrity of the RNA was checked by denaturing gel electrophoresis and concentration was determined as above. Total RNA was used for all PCR and cloning steps. Single oocytes were quickly homogenized in 100 µl of Tri-reagent using disposable plastic pestles. Thereafter, the supplier's protocol was followed. The centrifugation times were increased for precipitation of nucleic acids for up to 1 h at 14 000 g if short RNAs were to be detected.

Non-radioactive detection of nucleic acid

Assessment of RNA from injected oocytes was done using denaturing formaldehyde agarose gels (1.2%). RNA from single oocytes was separated and blotted by capillary force onto nylon membranes (Roche). After UV cross-linking the membranes were pre hybridized (>1 h) and hybridized over night in digoxygenin (DIG) EasyHyb solution (Roche). The DIG labelled probes were generated by *in vitro* transcription using the mMESSAGE mMACHINE kit (Ambion) and DIG labelled nucleotides (Roche). Hybridization was carried out at 58°C. Membranes were washed at a final stringency of 0.5× SSC/0.1% SDS at 50°C followed by ECL detection (Roche). To assay short RNAs samples were mixed at 1:3 ratio with Ambion formaldehyde loading buffer and heated at 75°C for 15 min. RNA was separated on 4–12% Bis–Tris polyacrylamide gels using 1× TBE running buffer [89 mM Tris-base, 2 mM EDTA-sodium salt, 89 mM boric acid, pH 8.5]. This combination gave superior result as compared to TBE-urea gels. Gels were pre-run for 1 h prior to loading of the samples. Blotting was performed at constant 100 mA for 50 min. All steps were performed using precast gels and a Criterion midi gel system (Biorad). Hybridization was essentially done as above, however, the stringency was reduced. Hybridization was performed at 37°C and final washes were at room temperature in 0.1× SSC, 0.1% SDS.

PCR and cloning

PCR-related techniques were used for expression analysis, cloning purposes and probe generation. To detect Na/phosphate transporter-related sense or antisense transcripts 0.1–0.5 µg of total RNA were reverse transcribed

using a kit from Qiagen. The RNA in 5 µl of water was denatured at 70°C for 5 min in the presence of 0.1 µg/µl oligo-dT primer and cooled to 37°C. 15 µl of RT mix was added containing buffer, nucleotides (250 µmol each), 4 units reverse transcriptase (Qiagen) and 20 units RNase inhibitor (Promega). After 60 min incubation at 37°C the supplier's protocol was downscaled and 2 µl of the RT reaction were used for hot start PCR with Taq polymerase and reagents from Qiagen. The final volume was 25 µl. For 3′ RACE an oligo dT adaptor primer (Sigma) was used to prime the reverse transcription. A first PCR was done as above including a gene-specific primer and the adaptor primer. A second, nested reaction was performed with 1 µl of the 1/2000 diluted first reaction, a second specific primer and the adaptor primer. PCR conditions were modified according to the annealing temperature of the primers and the length of the expected fragments. For 5′ RACE a kit from Invitrogen was used and the protocol was strictly followed. The sequences of the primers are given in the Supplementary Table 1. The two DNA fragments from 3′ and 5′ RACE were separated on agarose gels and purified. The entire fragment was amplified by overlapping PCR. One microlitre of each fragment was added to a PCR reaction (Qiagen). After four cycles at a low annealing temperatures (50°C) to allow hybridization of the overlapping DNA ends and extension of the full fragment the two adaptor primers were added and the cDNA was amplified. The resulting fragment was cloned into pCR2.1-TOPO using the TOPO cloning kit from Invitrogen according to the standard protocol. The cDNA was verified by sequencing.

Bisulfite treatment

Genomic DNA was isolated from adult mouse kidney, testis and skeletal muscle. The tissue was homogenized as above but omitting the step using the 25 G needle. The QIAamp DNA Mini Kit (Qiagen) was used according to the supplier's protocol. Concentration was determined by spectrophotometry. Approximately 0.5 µg of DNA was bisulfite treated using the EZ DNA methylation gold kit (Zymo Research). Colum purification of the DNA resulted in samples of 14 µl of which 3 µl were used for PCR in a 15 µl reaction (HotStarTaq Master Mix, Qiagen). The sequence of the primers specific for the sense and the antisense promoter as well as the sense/antisense overlapping region are given in the Supplementary Data. The resulting fragments were directly sequenced and checked for evidence of C/T bias in the context of CpG dinucleotides.

Bioinformatics

Probes from Affymetrix 430_2 arrays were mapped to the mouse genome using BLAT. The probes were correlated with exons and selected if they mapped to an exon on the opposite strand. The pipeline used the Affymetrix Mouse430_2 probe mapping generated in Ensembl v49 (http://Mar2008.archive.ensembl.org). Briefly, the Ensembl mapping pipeline retains genomic mappings with one mismatch regardless of the location on the genome, or multiple mappings. The probe mapping was done using exonerate (22). The antisense-specific probe coordinates represent about 3.5% of the total probes mapped to the genome. (The antisense probe sets are listed in Supplementary Table 2). The resulting probe set was used to assess datasets in the public repositories with emphasis on testis and embryonic stem cells. Other mouse tissues were included as controls and cross reference. The relevant GEO accession numbers are: GSE9954, different mouse tissues; GSE4193 spermatogenesis. Raw data was downloaded pre-processed with RMA and a per-gene normalisation applied in GeneSpring 7.3.1 (Agilent). Net expression levels were plotted for each probe and compared between the different tissues.

In order to assess a correlation between antisense transcription and random imprinting three published datasets were compared: Human antisense transcripts (16) were correlated with two compilations of monoallelically expressed human genes (23,24). HUGO gene names were used as common denominators for the three datasets. Genes were then identified that were present in both the antisense dataset and the imprinted genes pool. To assess significance of the overlap a random gene list of the same size as the imprinted gene set was compared to the antisense dataset over 100 000 permutations.

RESULTS

We compared the genomic organization of flounder and zebrafish *Slc34a* with the homologous gene encoding the Na/phosphate cotransporter *Slc34a1* from mouse. The presence of an annotated gene *Pfn3*, encoding the protein Profilin 3, closely downstream of the *Slc34a1* 3′end strongly indicated overlapping transcription of the mouse locus (25). An additional gene encoding the coagulation factor F12 (*F12*) could lead to overlapping transcripts as well (Figure 1). We tested for the expression of antisense transcripts by RT-PCR using primers located in different exons of *Slc34a1*, *Pfn3* and *F12*. In *Slc34a1* we focused on exons X and XIII based on structural information from the homologous zebrafish gene. Primers located in *Slc34a1* exon X and the in the distal part of *Pfn3* (in relation to *Slc34a1*) amplified specific fragments. Primers located in *Slc34a* exon XIII gave weak and unreliable signals and primers in the proximal part of *Pfn3* and *F12* failed to amplify any DNA (Figure 2A and not shown). RNA was extracted from different mouse tissues and tested for integrity by gel electrophoresis. cDNA was amplified from testis and kidney RNA. Skeletal muscle did not express the antisense transcript (Figure 2B).

We cloned the *Slc34a*-related antisense transcript by RACE and overlapping PCR from mouse testis RNA. The predominant transcript contained 517 nucleotides with an open reading frame of 158 amino acids. The first 47 amino acids are identical with Profilin 3, the remaining 111 amino acids show no similarity to known proteins and lack structural motifs. Other splice forms were identified, based on the inefficient amplification of the corresponding fragments we assume that the additional isoforms represent splice byproducts. These results

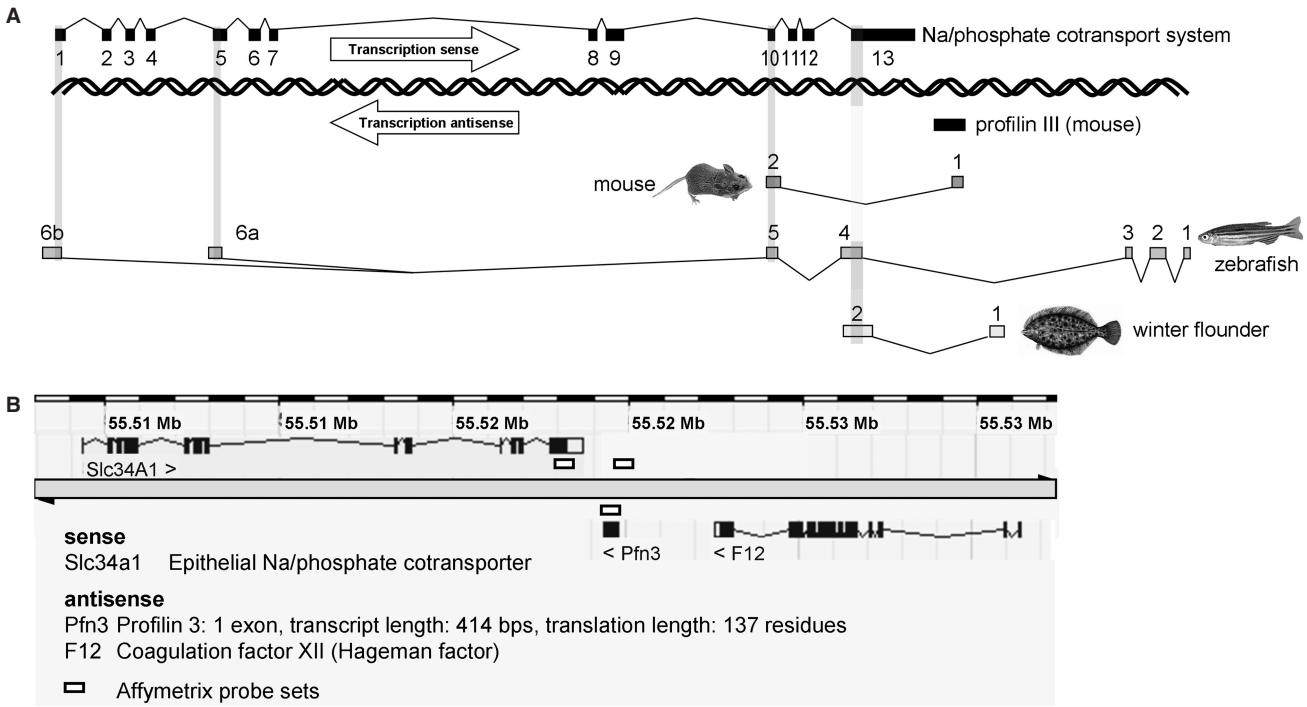

Figure 1. Schematic representation of the *Slc34a1* gene in vertebrates. (**A**) Phylogenetic conservation of the corresponding transcripts. The organization of the sense transcript (top) is conserved in all vertebrates with 13 exons and 12 introns. Antisense transcripts have been cloned from the three species flounder, zebrafish and mouse. The intron/exon structure of the different antisense transcripts differs significantly between organisms. Other splice forms may be expressed at a low level and have not been cloned. (**B**) The *Slc34a1* locus as it is displayed in Contig view in the public ensemble database. The Affymetrix probe sets recognize the 3′ end of the *Slc34a1* transcript (1423279_at), a region downstream of *Slc34a1* partly complementary to *Pfn3* (1427627_at) and the *Pfn3* transcript itself (1453962_at).

indicate that evolutionary conservation of the antisense transcript does not follow rules for protein coding RNAs. Open reading frames may be present; however, little pressure seems to act on maintaining them.

The significant expression of the *Slc34a*-related antisense transcript in both mice and zebrafish testis (26) raised the question if this was a gene-specific feature or if antisense transcription was generally unregulated in testis (27). Tools to routinely assess antisense transcripts on a genome wide scale are not established. We took advantage of a small but representative number of Affymetrix probes that recognize the reverse strand of coding genes. The probes were selected to localize in exonic regions of annotated sense transcripts. This enabled us to screen specifically for a limited number of transcripts that fulfill hallmarks of NATs as outlined in the introduction. On the Affymetrix mouse 430_2 array, 14 101 probes map to exons predicted by Ensembl but on the opposite strand. Among those probes 10 420 mapped to the genome once. The antisense probes represent 2596 unique exons of 1977 different genes (Ensembl v49). The individual probes belong to 1630 probe sets according to the standard Affymetrix CDF file. In other words, the 1630 probe sets return a positive expression call if (i) a potential antisense transcript is complementary to this exon and (ii) the antisense exon covers enough individual probes to return an unambiguous call (this is not necessarily the case because the intron–exon structure refers to the sense transcript. The complementary exon(s) of the antisense transcript may only span a fraction of the individual probes that make up the entire probe set). We tested the expression of the corresponding transcripts in various mouse tissues using GEO datasets GSE9954 and GSE4193. In particular, antisense transcripts were assessed in a multi tissue experiment including testis, seminal vesicle, brain, kidney and embryonic stem cells. Datasets from spermatocytes at different stages of development were included. We found significant up regulation of antisense transcripts in testis and, to a lesser extend, in seminal vesicle. Increased NATs expression was observed specifically during the final stages of spermatogenesis (round stage, Table 1). Interestingly, kidney and ES cells also showed a slightly increased expression of NATs, whereas values from other tissues were in a comparable low range.

We have previously reported that co-expressed overlapping *Slc34a* sense-antisense transcripts from zebrafish are processed into endo-siRNAs after injection into the nucleus of *Xenopus* oocytes (19). We used a similar approach to demonstrate that the mouse NAT was processed accordingly. As shown in Figure 3A single transcripts remained stable in both nucleus and cytoplasm. Co-expression of sense and antisense cRNAs in the nucleus resulted in complete degradation of both transcripts, whereas the related signals remained detectable after cytoplasmic injections (Figure 3B, 'full overlap').

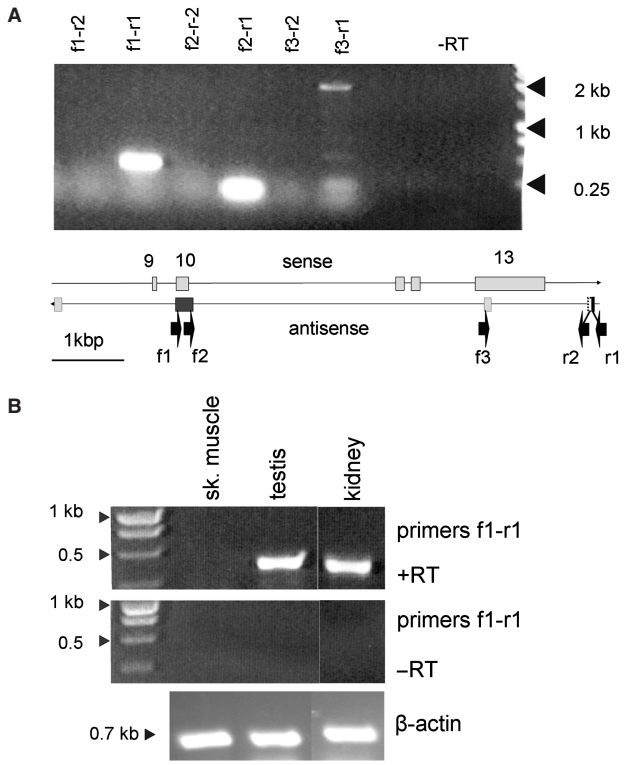

Figure 2. Analysis of the mouse *Slc34a1* antisense transcript. (**A**) Putative splice products were tested by RT–PCR using testis total RNA. Only primers located within sense exon 10 and at the profilin 5′end (f1-r1; f2-r1) gave detectable amplicons. The 2 kb band (f3-r1) derived from genomic DNA. If genomic DNA was used instead of cDNA long range PCR yielded the expected fragments of 2 kb (f3-r2) and about 6.5 kb (f1-r2 and f2-r2, data not shown). (**B**) Tissue distribution of the antisense transcript in mouse testis, kidney and skeletal muscle. The lower panels represent the negative control minus reverse transcriptase and β-actin, respectively. The locations of the primers are indicated on the scheme in (A).

The mouse sense and antisense transcripts share a single exon with an overlap of 124 bp. We determined the minimal size of overlap that induced processing by gradually truncating the antisense transcript to 40, 31, 30, 29 and 20 bp of complementarity. All the transcripts were stable for days after co-injection into the cytoplasm (C) of *Xenopus* oocytes (Figure 3B). Upon co expression of sense-antisense pairs in the nucleus (N) RNA overlaps of 30 bp or longer were efficiently cleaved. Shorter duplexes were either poorly processed (29 bp) or were stable (20 bp) (Figures 3 and 4). The generation of endo-siRNAs from overlapping sense/antisense transcripts was demonstrated by RNA PAGE and northern blotting of samples from single oocytes. In accordance with conventional northern blots short RNAs of approximately 23 bases were detectable in samples that showed evidence of sense/antisense processing (Figures 3 and 4, left panel lanes 1–5).

If the processing observed in *Xenopus* oocytes was physiologically relevant one would expect the presence of endo-siRNAs in tissues that co express *Slc34a1* sense and antisense transcripts [Figure 2B and (25)]. Total RNA samples from mouse kidney, testis and skeletal muscle were included into the PAGE northern experiments (Figure 4, left panel, lanes 6–8). As controls, samples from zebrafish embryos were included (Figure 4, right panel, lanes 9–11). Small RNAs are indeed present in both kidney and testis samples. Most interestingly, in testis both siRNA strands could be detected whereas in kidney only endo-siRNAs complementary to the antisense RNA were found. These results concur with published data from zebrafish *Slc34a* and other reports that showed accumulation of strand-specific endo-siRNAs (19–21). We concluded that strand selection of the endo-siRNAs was non-random. In the context of

Table 1. Expression of NATs in mouse tissues

Tissue	ES cells	Brain	Kidney	Testis	Sem. vesicle	Sem. v. curated[a]	Ovary	Placenta
Tissue distribution of antisense transcripts (NATs)								
Max value	149.8	70.85	188.3	254.9	1843	128.1	80.13	22.26
Min value	0.049	0.043	0.234	0.059	0.09	0.09	0.11	0.21
Mean	1.56	1.28	1.55	2.34	2.45	1.32	1.05	1.09
SD	4.47	2.91	7.04	10.08	45.73	3.19	2.21	1.24
Pfn3	1.07	0.77	1.06	18.75	1.39		0.66	0.80
Slc34a1	0.94	0.82	274.6	1.09	1.47		0.74	0.96
Slc34a-3′	0.99	0.75	49.49	1.46	1.62		0.72	0.84
Spermatogenesis								
Cell	A		B		Pachytene		Round	
Max value	4.07		3.50		5.30		24.42	
Min value	0.01		0.08		0.10		0.11	
Mean	0.98		1.00		1.11		1.21	
SD	0.33		0.32		0.41		1.38	
Pfn3	0.33		0.36		1.54		105.1	
Slc34a1	0.97		0.79		1.04		1.06	
Slc34a-3′	0.98		0.95		1.03		1.04	

Datasets GSE9954 and GSE4193 were evaluated using the probe sets specific for NATs (Supplementary data 1). The normalized values for the probesets *Pfn3* (1453962_at), *Slc34a1* (1423279_at) and *Slc34a-3′* (1427627_at) are included. The alternative splice product of *Pfn3* is not represented in the database; therefore probeset *Pfn3* (1453962_at) is included in the NATs probe compilation. The RT-PCR results indicate significant expression of the alternative splice form (Figure 2), however, the probeset will pick up the unspliced form as well.

[a]The curated data from seminal vesicle lacks the value of a single probeset 1422515_at: the probes bind to a highly expressed protein coding transcript (Svs7) that is specifically expressed in seminal vesicle and complementary to another protein coding transcript (AC113081).

siRNA-induced gene silencing this implies that a downstream response can be directed either towards the sense or the antisense strand. We therefore hypothesized a connection between the expression of NATs and epigenetic gene silencing, for example random monoallelic expression.

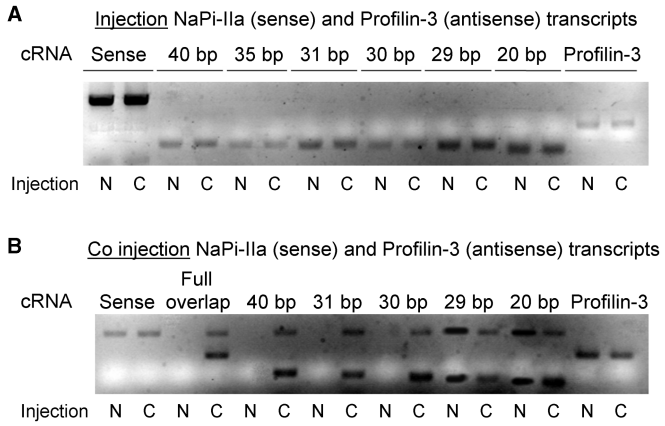

Figure 3. Expression of sense and antisense transcripts in *Xenopus* oocytes. (**A**) Injection of the different constructs into the cytoplasm (C) or nucleus (N) of oocytes. All the cRNAs remain stable regardless of cytoplasmic or nuclear injection. (**B**) Co-injection of sense and various truncated antisense constructs. The size of overlap is indicated. A 30 bp overlap is efficiently processed whereas a 29 bp overlap is relatively stable.

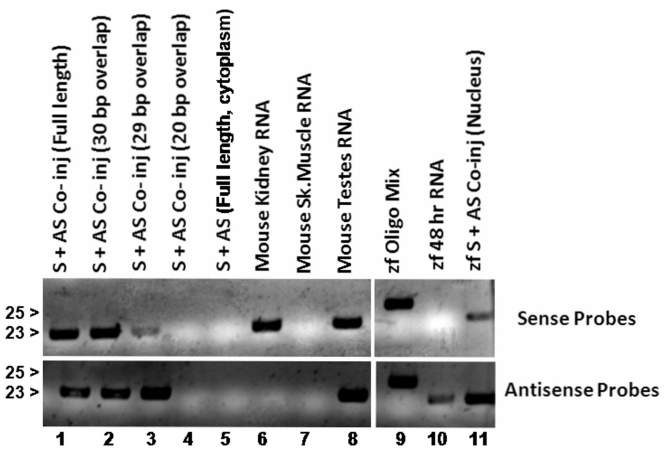

Figure 4. Northern blot analysis of short RNAs from injected oocytes and mouse tissues. The left panels (lanes 1–8) show samples from mouse, the right panel control samples from zebrafish (lanes 9–11). Lanes 1–3 indicates sense-antisense processing whereas lanes 4 and 5 do not contain short RNAs either because the overlap is too short (lane 4) or the samples were injected into the cytoplasm (lane 5). Lanes 6–8 represent tissues that express the sense encoded protein (kidney), do not express the transporter (skeletal muscle) or expresses both sense and antisense RNAs but the presence of the transporter is unclear (testis). Lane 10 represents short RNAs isolated from 48 h zebrafish embryos (19) as a positive control. Lane 11 shows another control, i.e. a sample from *Xenopus* oocytes injected into the nucleus with sense and antisense RNA. 'Sense'- and 'antisense'- probes mean that the short RNAs detected with the sense probe will be complementary to the antisense transcript and *vice versa*.

Random monollelic expression is experimentally approached by using polymorphisms to assess the allelic origin of transcripts. We argued that polymorphisms or mutations within a bi-directionally transcribed locus may skew or completely reverse the endo-siRNA strand selection process. This would result in reversed targeting of endo-siRNA-induced silencing. The sense transcript instead of the antisense transcript would become silenced on the allele that shows the polymorphism. The other allele, unaffected by the polymorphism, would show the 'normal' strand selection. As a consequence, mono-allelic expression of the polymorphic gene locus should be observed. To support this hypothesis we investigated whether a genome wide correlation between mono-allelic gene expression (i.e. random imprinting) and antisense transcription could be found. Both phenomena have been assessed in humans by other groups and the corresponding datasets are published (16,24). We used the datasets from Chen *et al.* for human NATs (2940) and Gimelbrant *et al.* for monoallelically expressed human autosomal genes (16,24). As a control group, the 25 157 HUGO genes were used. Matching the gene names of the three data sets resulted in 2772 NATs and 281 randomly imprinted genes with compatible annotations. The overlap between the NATs dataset and monoallelically expressed genes was 56, a figure in the 99th percentile when compared against the control of a randomly permutated list of 281 HUGO genes and the NATs. A similar evaluation was performed with an independent set of randomly imprinted genes (23) and (11) with comparable results (Table 2). These findings add weight to the argument that mutations may influence the orientation of endo-siRNAs.

To assess for short RNA-induced transcriptional silencing we checked for DNA methylation changes in the sense/antisense overlapping region (5 CpG dinucleotides in 274 bp) and the sense promoter (1 CpG in 374 bp) in DNA isolated from mouse kidney and muscle, respectively. These regions are devoid of CpG islands. Bisulfite treatment of the DNA and direct sequencing of the amplified fragments revealed complete CpG methylation and no difference between kidney and muscle. Therefore gene regulation by NATs may include other mechanisms such as histone modifications (28) to induce transcriptional silencing.

Table 2. Correlation between NATs and randomly imprinted genes

Monoalleically expressed genes	Antisense transcripts	HUGO genes	Overlap	Significance (%)
104 (23)	2772 (16)		23	>95
	2772 (16)	104 from 25 157	11 ± 3	
281 (24)	2772 (16)		56	>99
	2772 (16)	281 from 25 157	30 ± 5	

Two datasets that identified randomly imprinted human genes were compared to the human antisense transcripts (16,23,24). As a control the same comparison was done with 281 and 104 randomly selected genes from the HUGO database (100 000 times).

DISCUSSION

Our experiments addressed the phenomenon of antisense transcription at single gene and entire genome level. We aimed at correlating the data and suggest novel strategies to assess gene regulation by NATs. It is important to take into account that the conclusions from this approach only apply to processed NATs that show exon complementary with the corresponding processed sense transcript.

We report the cloning of a NAT related to the Na/phosphate cotransporter gene *Slc34a1* from mouse testis. The NAT represents an alternative splice product of *Pfn3* that has not been reported previously in the database (EU375560). Our approach exemplifies the difficulties related to the cloning of a specific NAT. The poor phylogenetic conservation of NATs (Figure 1) and their often unreliable representation in the public databases complicate any cloning effort of a specific NAT. We were fortunate as the *Pfn3* gene provided a promising starting point for a RACE approach. In addition, the *Slc34a1* NAT contained an intron allowing for unambiguous annotation of PCR products. Without an intron, the confirmation of antisense orientation can be cumbersome and would rely on more error prone strategies such as strand-specific RT, oligo-dT priming, capping or transcription start site predictions. The generally low expression level and the restricted expression pattern of NATs add further complications to a targeted cloning approach. The general up regulation of NATs in testis [Table 1 and (27)] would now suggest using this tissue in relation with a NAT cloning effort. In addition, investigations into comprehensive antisense expression and transcript characterization may best focus on the testis transcriptome.

The expression pattern of NATs in testis is particularly intriguing with respect to the developmental stage at which overexpression occurs. During the early polyploid stages, types A, B and pachytene (Table 1), expression of NATs is low. It significantly increases during later stages of spermiogenesis in haploid cells. Widespread translational repression is observed during this later stage of spermiogenesis (29).

We found a rather insignificant expression of NATs in other organs apart from testis whereas other reports find low but significant expression of NATs in other tissues. This discrepancy is most likely due to experimental constraints. We used stringent conditions when analyzing the DNA array data and may have excluded weak signals that were included in other reports (30,31). Transcripts specific to a very small number of cells in a heterogeneous tissue may have failed our detection limit. More sensitive experiments such as RT–PCR indeed confirmed the co expression of the complementary *Slc34a* transcripts [Figure 2 and (25,26)].

The cloned *Slc34a1* NAT is processed into mRNA and displays exon complementarity with the corresponding sense transcript. There is compelling recent evidence that NAT/sense transcript hybrids may be cleaved into endo-siRNAs (19,21,32,33). The detection of *Slc34a1*-related endo-siRNAs in testis concurs with these previous findings (19,27). We also found that a 30 bp overlap of the two substrate RNAs represents the cut off for efficient processing (Figure 3B). Dicer accepts long double-stranded RNAs as substrates and produces short RNAs of variable size [19–30 bp (34)]. A significant decrease in dicer processivity between 30 and 29 bp double-stranded substrates was also reported by Elbashir *et al*. (35). These findings suggest that the short *Slc34a1*-related RNAs are indeed dicer processed endo-siRNAs. The tissue-specific orientation of the endo-siRNAs in testis and kidney (Figure 4) represents an exciting finding that corroborates previous observations in mouse and zebrafish. Watanabe *et al*. reported endo-siRNAs related to the *Ppp4r1* gene that were all oriented in sense direction in mouse oocytes, thus reflecting a similar situation as reported for the *Slc34a1* endo-siRNAs in mouse kidney. In zebrafish embryos, the orientation of *Slc34a* endo-siRNAs shows developmental regulation from antisense direction to sense orientation. Interestingly, the switch coincides with the physiological induction of the sense-encoded protein. These observations suggest that the orientation of the endo-siRNAs is somehow related to the expression of the sense transcript. The fact that northern blotting revealed different endo-siRNA patterns in testis and kidney (Figure 4) might relate to the transcriptional state of the relevant gene locus. In testis, a general up regulation of transcription would stimulate sense and antisense transcription at a similar scale leading to transient levels of both endo-siRNA strands (36). In kidney, the *Slc34a1* sense transcript is activated in a gene-specific way. This could lead to either a short period of sense/antisense co-expression followed by sustained sense expression or, alternatively, strong sense expression including stochastic antisense events at a low level. In both cases, it is conceivable that only the sense oriented endo-siRNAs (targeting antisense transcription) are predominantly accumulated.

By comparing published datasets we found a significant correlation between NATs expression and monoallelically transcribed genes (16,23,24). The data sets, however, contain monoallelically expressed genes without NATs and *vice versa*. A better overlap between the datasets is likely to be achieved once the NAT transcriptome becomes systematically characterized. Random monoallelic expression of a gene is assessed using polymorphisms to determine the allelic origin of transcripts. On the other hand, we have shown that the probability of monoallelic gene expression increases with the expression of NATs and possibly endo-siRNAs. It is therefore tempting to speculate that the expression of NATs from a polymorphous locus, endo-siRNA strand selection and random allelic expression are connected phenomena. Such a link could, for example, explain the selection against NATs on the mammalian X chromosome because a NAT-induced silencing event would lead to a complete knock down of the relevant gene. In this context the identification of the factors that influence endo-siRNA strand selection would be of major interest (37).

We developed this idea further towards a concept which is outlined in Figure 5. The working model predicts that sense and antisense transcripts are co expressed during a limited window during cell development or differentiation. Our expression studies indicate that such co-expression

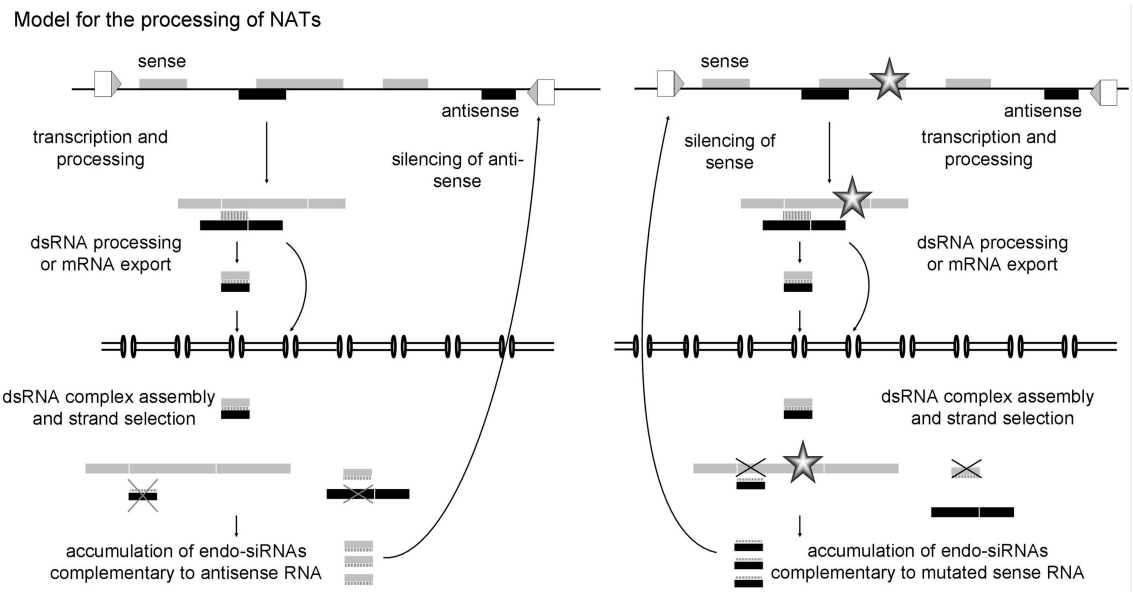

Figure 5. Model for the biological role of antisense transcripts. (**A**) The 'normal' case where the antisense driving promoter becomes silenced. (**B**) The situation of a mutated allele (red star). The discontinuous nature of transcription may influence the timing of promoter silencing. This could ensure that the feedback mechanism remains allele specific.

could occur during late stages of spermiogenesis or in renal stem cells (Table 1). Production of endo-siRNA and possibly the concomitant accumulation of specifically oriented endo-siRNA strands involve fully processed mRNAs. Experimental evidence so far (19,21) predicts that the protein coding sense transcript dictates the strand selection. Features that differentiate sense and antisense transcripts and may affect endo-siRNA strand selection include, translatability (38), structure (39) or stability, for example. It is conceivable that polymorphisms or mutations in general alter these transcript features and skew endo-siRNA strand selection. As a consequence, the silencing response could become redirected and target the sense transcript (40–42,43). The entire process could be seen as a quality test for RNAs.

The proposed concept has different consequences if applied to the tissues testis and kidney that show increased NATs expression. In testis, male germ cells show significant up regulation of transcriptional activity including the expression of NATs (27,44). This will lead to a wide spread generation of endo-siRNAs, hence both endo-siRNA strands are detected on the northern blot in Figure 4. The concomitant RNA quality test will lead to the accumulation of sense oriented endo-siRNAs (complementary to the antisense transcript; quality test passes). If a specific gene is mutated the quality control potentially fails, endo-siRNA strand selection is reversed and the sense transcript becomes silenced. The haploid spermatids will not be able to compensate for the loss and face increased selective pressure. Significant levels of apoptosis are indeed detected during spermatogenesis (45). The process may therefore remove cells that carry deleterious mutations and eventually result in a purified population of spermatozoa.

In kidney, the RNA quality control would apply as above, however with slightly different consequences. Under physiological conditions, the endo-siRNA complementary to the antisense RNA is selected and the sense transcript is expressed. If one allele is mutated and shows reversed strand selection only one allele would become silenced and the other one stays active. Allele specificity may be achieved simply by the timing of transcription from the two alleles. The specific cell may still function normally but show 'random' monoallelic expression of the corresponding gene.

To conclude, we suggest that NATs are a significant source of endo-siRNAs. We hypothesize that these endo-siRNAs act as molecular switches responsive to mutations/polymorphisms. Such a biological role may explain the under representation of NATs on mammalian X-chromosomes and the phenomenon of mono-allelic gene expression.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

FUNDING

Newcastle University; Biotechnology and Biological Sciences Research Council UK studentship [to M.C.]. Funding for open access charge: Newcastle University.

Conflict of interest statement. None declared.

REFERENCES

1. Beiter,T., Reich,E., Williams,R.W. and Simon,P. (2008) Antisense transcription: A critical look in both directions. *Cell Mol. Life Sci.*, **66**, 94–112.
2. Chen,J., Sun,M., Rowley,J.D. and Hurst,L.D. (2005) The small introns of antisense genes are better explained by selection for rapid transcription than by "genomic design". *Genetics*, **171**, 2151–2155.

3. Baulcombe,D. (2004) RNA silencing in plants. *Nature*, **431**, 356–363.
4. Wagner,E.G. and Simons,R.W. (1994) Antisense RNA control in bacteria, phages, and plasmids. *Annual Rev Microbiol.*, **48**, 713–742.
5. Lapidot,M. and Pilpel,Y. (2006) Genome-wide natural antisense transcription: coupling its regulation to its different regulatory mechanisms. *EMBO Rep.*, **7**, 1216–1222.
6. Lavorgna,G., Dahary,D., Lehner,B., Sorek,R., Sanderson,C.M. and Casari,G. (2004) In search of antisense. *Trends Biochem. Sci.*, **29**, 88–94.
7. Werner,A. and Berdal,A. (2005) Natural antisense transcripts: sound or silence? *Physiol. Genomics*, **23**, 125–131.
8. Goldmit,M. and Bergman,Y. (2004) Monoallelic gene expression: a repertoire of recurrent themes. *Immunol. Rev.*, **200**, 197–214.
9. Yang,P.K. and Kuroda,M.I. (2007) Noncoding RNAs and intranuclear positioning in monoallelic gene expression. *Cell*, **128**, 777–786.
10. Katayama,S., Tomaru,Y., Kasukawa,T., Waki,K., Nakanishi,M., Nakamura,M., Nishida,H., Yap,C.C., Suzuki,M., Kawai,J. *et al.* (2005) Antisense transcription in the mammalian transcriptome. *Science*, **309**, 1564–1566.
11. Werner,A., Carlile,M. and Swan,D. (2009) What do natural antisense transcripts regulate? RNA Biol., **6**, 43–48.
12. Zhang,Q., Zhang,J., Moe,O.W. and Hsia,C.C. (2008) Synergistic upregulation of erythropoietin receptor (EPO-R) expression by sense and antisense EPO-R transcripts in the canine lung. *Proc. Natl Acad. Sci. USA*, **105**, 7612–7617.
13. Faghihi,M.A., Modarresi,F., Khalil,A.M., Wood,D.E., Sahagan,B.G., Morgan,T.E., Finch,C.E., St Laurent,G. 3rd, Kenny,P.J. and Wahlestedt,C. (2008) Expression of a noncoding RNA is elevated in Alzheimer's disease and drives rapid feed-forward regulation of beta-secretase. *Nat. Med.*, **14**, 723–730.
14. Coudert,A.E., Pibouin,L., Vi-Fane,B., Thomas,B.L., Macdougall,M., Choudhury,A., Robert,B., Sharpe,P.T., Berdal,A. and Lezot,F. (2005) Expression and regulation of the Msx1 natural antisense transcript during development. *Nucleic Acids Res.*, **33**, 5208–5218.
15. Hastings,M.L., Ingle,H.A., Lazar,M.A. and Munroe,S.H. (2000) Post-transcriptional regulation of thyroid hormone receptor expression by cis-acting sequences and a naturally occurring antisense RNA. *J. Biol. Chem.*, **275**, 11507–11513.
16. Chen,J., Sun,M., Kent,W.J., Huang,X., Xie,H., Wang,W., Zhou,G., Shi,R.Z. and Rowley,J.D. (2004) Over 20% of human transcripts might form sense-antisense pairs. *Nucleic Acids Res.*, **32**, 4812–4820.
17. Yelin,R., Dahary,D., Sorek,R., Levanon,E.Y., Goldstein,O., Shoshan,A., Diber,A., Biton,S., Tamir,Y., Khosravi,R. *et al.* (2003) Widespread occurrence of antisense transcription in the human genome. *Nat. Biotechnol.*, **21**, 379–386.
18. Kiyosawa,H., Yamanaka,I., Osato,N., Kondo,S. and Hayashizaki,Y. (2003) Antisense transcripts with FANTOM2 clone set and their implications for gene regulation. *Genome Res.*, **13**, 1324–1334.
19. Carlile,M., Nalbant,P., Preston-Fayers,K., McHaffie,G.S. and Werner,A. (2008) Processing of naturally occurring sense/antisense transcripts of the vertebrate Slc34a gene into short RNAs. *Physiol. Genomics*, **34**, 95–100.
20. Tam,O.H., Aravin,A.A., Stein,P., Girard,A., Murchison,E.P., Cheloufi,S., Hodges,E., Anger,M., Sachidanandam,R., Schultz,R.M. *et al.* (2008) Pseudogene-derived small interfering RNAs regulate gene expression in mouse oocytes. *Nature*, **453**, 534–538.
21. Watanabe,T., Totoki,Y., Toyoda,A., Kaneda,M., Kuramochi-Miyagawa,S., Obata,Y., Chiba,H., Kohara,Y., Kono,T., Nakano,T. *et al.* (2008) Endogenous siRNAs from naturally formed dsRNAs regulate transcripts in mouse oocytes. *Nature*, **453**, 539–543.
22. Slater,G.S. and Birney,E. (2005) Automated generation of heuristics for biological sequence comparison. *BMC Bioinformatics*, **6**, 31.
23. Yang,H.H., Hu,Y., Edmonson,M., Buetow,K. and Lee,M.P. (2003) Computation method to identify differential allelic gene expression and novel imprinted genes. *Bioinformatics*, **19**, 952–955.
24. Gimelbrant,A., Hutchinson,J.N., Thompson,B.R. and Chess,A. (2007) Widespread monoallelic expression on human autosomes. *Science*, **318**, 1136–1140.
25. Braun,A., Aszodi,A., Hellebrand,H., Berna,A., Fassler,R. and Brandau,O. (2002) Genomic organization of profilin-III and evidence for a transcript expressed exclusively in testis. *Gene*, **283**, 219–225.
26. Nalbant,P., Boehmer,C., Dehmelt,L., Wehner,F. and Werner,A. (1999) Functional characterization of a Na+-phosphate cotransporter (NaPi-II) from zebrafish and identification of related transcripts. *J. Physiol.*, **520 (Pt 1)**, 79–89.
27. Okada,Y., Tashiro,C., Numata,K., Watanabe,K., Nakaoka,H., Yamamoto,N., Okubo,K., Ikeda,R., Saito,R., Kanai,A. *et al.* (2008) Comparative expression analysis uncovers novel features of endogenous antisense transcription. *Hum. Mol. Genet.*, **17**, 1631–1640.
28. Janowski,B.A., Younger,S.T., Hardy,D.B., Ram,R., Huffman,K.E. and Corey,D.R. (2007) Activating gene expression in mammalian cells with promoter-targeted duplex RNAs. *Nat. Chem. Biol.*, **3**, 166–173.
29. Steger,K. (1999) Transcriptional and translational regulation of gene expression in haploid spermatids. *Anat. Embryol.*, **199**, 471–487.
30. Oeder,S., Mages,J., Flicek,P. and Lang,R. (2007) Uncovering information on expression of natural antisense transcripts in Affymetrix MOE430 datasets. *BMC Genomics*, **8**, 200.
31. Werner,A., Schmutzler,G., Carlile,M., Miles,C.G. and Peters,H. (2007) Expression profiling of antisense transcripts on DNA arrays. *Physiol. Genomics*, **28**, 294–300.
32. Czech,B., Malone,C.D., Zhou,R., Stark,A., Schlingeheyde,C., Dus,M., Perrimon,N., Kellis,M., Wohlschlegel,J.A., Sachidanandam,R. *et al.* (2008) An endogenous small interfering RNA pathway in Drosophila. *Nature*, **453**, 798–802.
33. Birchler,J.A. and Kavi,H.H. (2008) Molecular biology. Slicing and dicing for small RNAs. *Science*, **320**, 1023–1024.
34. MacRae,I.J., Zhou,K. and Doudna,J.A. (2007) Structural determinants of RNA recognition and cleavage by Dicer. *Nat. Struct. Mol. Biol.*, **14**, 934–940.
35. Elbashir,S.M., Harborth,J., Lendeckel,W., Yalcin,A., Weber,K. and Tuschl,T. (2001) Duplexes of 21-nucleotide RNAs mediate RNA interference in cultured mammalian cells. *Nature*, **411**, 494–498.
36. Shima,J.E., McLean,D.J., McCarrey,J.R. and Griswold,M.D. (2004) The murine testicular transcriptome: characterizing gene expression in the testis during the progression of spermatogenesis. *Biol. Reprod.*, **71**, 319–330.
37. Krueger,C. and Morison,I.M. (2008) Random monoallelic expression: making a choice. *Trends Genet.*, **24**, 257–259.
38. Zhou,R., Hotta,I., Denli,A.M., Hong,P., Perrimon,N. and Hannon,G.J. (2008) Comparative analysis of argonaute-dependent small RNA pathways in Drosophila. *Mol. Cell*, **32**, 592–599.
39. Rudnick,S.I., Swaminathan,J., Sumaroka,M., Liebhaber,S. and Gewirtz,A.M. (2008) Effects of local mRNA structure on post-transcriptional gene silencing. *Proc. Natl Acad. Sci. USA.*, **105**, 13787–13792.
40. Iida,T., Nakayama,J. and Moazed,D. (2008) siRNA-mediated heterochromatin establishment requires HP1 and is associated with antisense transcription. *Mol. Cell*, **31**, 178–189.
41. Morris,K.V., Chan,S.W., Jacobsen,S.E. and Looney,D.J. (2004) Small interfering RNA-induced transcriptional gene silencing in human cells. *Science*, **305**, 1289–1292.
42. Morris,K.V., Santoso,S., Turner,A.M., Pastori,C. and Hawkins,P.G. (2008) Bidirectional transcription directs both transcriptional gene activation and suppression in human cells. *PLoS Genet.*, **4**, e1000258.
43. Verdel,A., Jia,S., Gerber,S., Sugiyama,T., Gygi,S., Grewal,S.I. and Moazed,D. (2004) RNAi-mediated targeting of heterochromatin by the RITS complex. *Science*, **303**, 672–676.
44. Kimmins,S., Kotaja,N., Davidson,I. and Sassone-Corsi,P. (2004) Testis-specific transcription mechanisms promoting male germ-cell differentiation. *Reproduction*, **128**, 5–12.
45. Baum,J.S., St George,J.P. and McCall,K. (2005) Programmed cell death in the germline. *Semin. Cell Dev. Biol.*, **16**, 245–259.