What are natural antisense transcripts good for?

Andreas Werner*1 and Daniel Swan†

*RNA Biology Group, Institute for Cell and Molecular Biosciences, Newcastle University, Framlington Place, Newcastle upon Tyne NE2 4HH, U.K., and †Bioinformatics Unit, Institute for Cell and Molecular Biosciences, Newcastle University, Framlington Place, Newcastle upon Tyne NE2 4HH, U.K.

Abstract

NATS (natural antisense transcripts) are important regulators of eukaryotic gene expression. Interference between the expression of protein-coding sense transcripts and the corresponding NAT is well documented. In the present review, we focus on an additional, higher-order role of NATs that is currently emerging. The recent discovery of endogenous siRNAs (short interfering RNAs), as well as NAT-induced transcriptional gene silencing, are key to the proposed novel function of NATs.

Introduction

The sequencing of the human genome revealed that the number of protein-coding genes does not scale with organismal complexity [1]. Interestingly, most of the coding genes are well-conserved in animal evolution and orthologous proteins fulfil similar biological roles in different species. The difference in biological complexity in animals is therefore unlikely to be entwined in protein-coding genes. Emerging evidence from several areas of research strongly supports the view that non-protein-coding RNA is essential in establishing organismal complexity [2]. Driven by rapid technical advances in transcriptome analysis and high-throughput sequencing, the overwhelming complexity of the non-protein-coding transcriptome is currently emerging. One of the formidable challenges will be to separate non-protein-coding transcripts with a regulatory role in gene expression from transcriptional noise. Considering the sheer amount of information that has recently been published in the field of non-protein-coding RNAs, this review attempts to not miss the forest for the trees, but also avoid cherry-picking.

Genome-wide transcription

Tiling arrays and high-throughput sequencing strategies have revealed a plethora of non-protein-coding transcripts from both genic and intergenic regions, indeed most of the genome seems to be sporadically transcribed [3]. Based on a 200 nucleotide cut-off according to RNA purification protocols, the transcripts are classified as short RNAs and long non-protein-coding RNAs [4]. Structural and functional features can be used to classify transcripts into broad groups with common biological roles or raisons d’être. A brief summary of selected RNA groups that are often discussed in the context of NATs (natural antisense transcripts) will be given. These include various classes of short RNAs, promoter-associated transcripts and long non-protein-coding RNAs (lincRNAs). The short profiles of the selected RNA groups are based on findings from animal model systems. This limitation is to avoid the dilemma of studying drivers of organismal complexity (non-protein-coding RNAs) in systems that score comparably low in that very discipline.

Short RNAs

Short RNAs derive from double-stranded RNA precursors by the processing of type III RNA endonucleases. The resulting oligonucleotides of 20–24 nucleotides promote post-transcriptional gene silencing. The two main classes of short RNAs are miRNAs (microRNAs) and siRNAs (short interfering RNAs) [5]. A third class of short RNAs, piRNAs (piwi-interacting RNAs) are slightly longer (26–32 nucleotides) and derive from single-stranded precursors [6].

miRNAs are important regulators of gene expression at the level of translation, with an essential input in developmental processes. Inhibitory complexes are guided to target mRNAs that are largely, but not fully, complementary to the miRNA. Inhibition is not necessarily linked to mRNA degradation [5].

siRNAs are perfectly complementary to the target sequence and result in endonucleic cleavage of the mRNA. In plants, siRNAs represent a powerful defence strategy against viruses. Accordingly, plant cells produce virus-derived siRNAs upon infection, whereas animal cells do not [7]. Defence against viruses is largely covered by the immune system and the biological role of siRNAs in animals is speculative [8]. siRNAs that derive from endogenous sources (endo-siRNAs) will be discussed in conjunction with NATs below.

piRNAs are germ-line-specific short RNAs, and male mice that lack key enzymes in piRNA processing become sterile [9]. In contrast with the other short RNAs mentioned so far, the biogenesis of piRNAs does not depend on RNase type III enzymes [10]. They derive from distinct non-coding

Key words: endogenous short interfering RNA; monoallelic gene expression; natural antisense transcript; non-protein-coding RNA; random imprinting; transcriptional silencing.

Abbreviations used: Ago, Argonaute; lincRNA, long non-protein-coding RNA; miRNA, microRNA; NAT, natural antisense transcript; PASR, promoter-associated short RNA; piRNA, piwi-interacting RNA; PIWIF, promoter upstream transcript; RISC, RNA-induced silencing complex; RNA interference; siRNA, short interfering RNA; endo-siRNA, endogenous siRNA; TAAR, tetrans-associated short RNA.

†To whom correspondence should be addressed (email andreas.werner@ncl.ac.uk).
regions of the genome and suppress transposon activity by transcriptional silencing [11].

**Promoter associated transcripts**

Pioneering studies by Kapranov and others [12,13] identified short RNAs and unstable transcripts of up to 600 nucleotides which mapped to gene boundaries. Several distinct RNA species were observed that originated either from transcription start sites [PASRs (promoter-associated short RNAs)], terminator sites [TASRs (termini-associated short RNAs)] or promoter regions [PROMPTs (promoter upstream transcripts)]. PASRs and TASRs are between 20 and 90 nucleotides long, whereas PROMPTs are usually longer than 200 nucleotides [13]. Interestingly, all of the different RNAs occur in sense and antisense orientations and are generally short lived [14]. The transcripts are assumed to reflect inherent properties of RNA polymerase II, such as promiscuous promoter binding and transcription initiation, as well as discontinuous elongation [15,16]. The fact that both gene boundaries display these spurious RNAs refers to a putative loop structure of transcribed genes, but also raises the intriguing perspective of synthesizing stable natural antisense transcripts [17].

**lincRNAs**

lincRNAs play a well-described role in X chromosome inactivation and parental imprinting. The transcripts Air, Kcnq1ot1 or Xist, for example, recruit repressive regulatory complexes and result in chromatin silencing of the corresponding locus or X allele respectively [18]. Another example of RNA-directed repressive chromatin modifications has recently been reported for the developmental regulation of Hox gene expression [19]. Interestingly, the non-protein-coding RNA essential to this process, called HOTAIR, is transcribed from an intergenic region and acts in trans. The idea of regulatory RNAs from intergenic regions was further developed: based on chromatin marks, over a thousand lincRNAs were identified and experimentally confirmed [20]. Many of these transcripts recruit chromatin-modifying complexes; however, other mechanisms of action also apply [18,21].

**NATs**

A particularly intriguing group of non-protein-coding RNAs are NATs. These are generally non-protein-coding, but fully processed, mRNAs that are transcribed from the opposite strand of protein-coding sense transcripts [22]. After splicing, the sense and the corresponding antisense transcripts share complementary exons and potentially form RNA–RNA hybrids. To highlight the potential regulatory impact on the cognate sense transcript, NATs are also referred to as cis–NATs. This definition of NATs originates from the pre-genomic area when the vast majority of non-protein-coding RNAs included the well-characterized transfer, ribosomal and spliceosomal RNAs, plus a few antisense transcripts from genic loci. Experiments involving NATs usually used the criterion of mRNA processing to distinguish between breakdown products of sense transcripts and ‘real’ NATs. These constraints applied to both large-scale cDNA sequencing projects pioneered by the FANTOM consortium as well as early bioinformatic approaches [23–25].

The focus on genic transcripts gives an incomplete account of cellular non-protein-coding and complementary RNAs; nevertheless, these early studies revealed a number of important findings. Despite the rather stringent limitations to qualify as a NAT (evidence of mRNA-like processing), up to 72% of all mouse genes show evidence of bi-directional transcription [26]. Comparable numbers are found in human and other animals. Caenorhabditis elegans represents an exception with a markedly decreased number of NATs [27]. Phylogenetic comparison of NATs in human, mouse and puffer fish indicated conservation of antisense transcription [28]. Neither coding potential nor splice structure, however, are under stringent selective pressure. The most striking feature of NATs represents their significant under-representation on the human and murine X chromosome [23,29]. Interestingly, NATs that are expressed from mammalian X chromosomes tend to be clustered in areas that escape inactivation [29]. In addition, flies and worms that adjust the output from the X chromosomes, but keep both alleles active, show equal distribution of NATs on the X chromosomes and autosomes [27]. These findings indicate that the expression of NATs is favourable on a bi-allelic background, but counter-indicated if monoallelically expressed. Interestingly, the bias only applies to NATs with exon complementarity to the sense transcript, but not antisense transcription as such. This observation indicates that an RNA–RNA intermediate is essential in the biological role of NATs. Taken together, these findings point towards a higher-order biological function of NATs that is not linked to the ontology of bi-directionally transcribed genes [30]. An emerging hypothesis describing this higher-order role will be presented below.

On the other hand, there is a wealth of well-documented examples of NATs that regulate the expression of their corresponding protein-coding sense transcripts [31]. Generally, increased levels of a specific NAT have an inhibitory influence on the sense transcript. However, NATs that protect the sense transcript against nuclease degradation have also been reported [32]. These findings are of clinical relevance because genes that are linked to cancer and other diseases are regulated by NATs. Excellent recent reviews discuss the gene-specific effects of NATs in detail, therefore this aspect will not feature prominently in the present review [33].

**A higher-order function of NATs**

Expression of NATs in specific tissues provides clues about their physiological role. Unfortunately, a comprehensive atlas of NATs is yet to be established. Genome-wide expression of NATs has been reported; however, the data were generated using human cell lines and is therefore of limited use to inquire tissue-specific expression patterns [34].
used Affymetrix gene expression arrays to assess the level of NATs in selected tissues [35]. The commercial arrays contain probe sets that hybridize to the opposite strand of annotated exons and will therefore bind to antisense transcripts complementary to this region [35–37]. We identified the probe sets and used them to mine datasets from different mouse tissues. NATs, as represented by the unbiased selection of probes, were most prominently expressed in testis, more specifically in haploid spermatids [38]. This finding was confirmed using other approaches [39]. Low signals were found in all tissues [25,35] and, interestingly, NATs were strongly correlated with the expression of the corresponding sense transcript. These studies do not specify whether sense and antisense transcripts are expressed in the same cell, but co-expression has been confirmed for specific bi-directionally transcribed genes [40]. Consequently, the hypothesized higher-order function for NATs is likely to be important during spermatogenesis, but may also be relevant in other organs or specific cell populations. In addition, the expression of NATs seems to be linked to the expression of the sense transcripts and involve the formation of RNA hybrids.

NATs as a source of endo-siRNAs

Long perfect RNA-hybrids in the cytoplasm of a cell are a sign of viral infection and trigger a stress response through protein kinase R. In the nucleus, however, RNA duplexes may have other functions. The sense/antisense hybrids could become edited by the enzyme Adar (adenosine deaminase acting on RNA) or alternatively feed into an RNAi (RNA interference)-related pathway [41]. A bioinformatic survey revealed that RNA editing was prevalent in introns of primary transcripts, but absent from sense/antisense complementary exons [42]. The connection between complementary transcripts and RNAi was initially challenged due to the lack of NAT-related clones in short RNA-sequencing projects. More recent RNA-Seq results and our own work, however, corroborate a link between NATs and RNAi [38,43–45]. The term ‘endo-siRNAs’ was coined to highlight the endogenous origin of these short RNA species. Most interestingly, only one strand of the original duplex was usually detected, indicating non-random processing of the double-stranded RNA precursor. We observed a switch in orientation of genespecific endo-siRNAs that was developmentally induced during zebrafish embryogenesis (Figure 1). The orientation correlated with the expression level of the corresponding sense or antisense transcript [43]. Detection of both strands of siRNAs and miRNAs confirms that strand selection of short RNAs may not be merely driven by the physico-chemical properties of the RNA hybrid, but also by the amount of sense or the antisense transcript [46–49].

NATs and transcriptional gene silencing

RNA-directed transcriptional gene silencing is well-established in yeast and plants. To what extent these findings were relevant to mice and humans was unclear because mammals lack one of the key enzymes of the process which enables amplification of the RNA signal [50]. Few endogenous antisense RNAs were found to silence the corresponding sense transcripts, indicating that a comparable mechanism may also apply in mammals. For example, a rare form of α-thalassaemia was shown to be caused by a rearranged, constitutively expressed LUC7 gene. The LUC7 mRNA is antisense to HBA2 and resulted in methylation and concomitant inactivation of the HBA2 promoter [51]. In addition, the tumour suppressor gene p15 was found to be repressed by a NAT in leukaemia [52]. The specific role of the (antisense) RNAs in these processes, however, was unclear. Pioneering work from K. Morris’s laboratory and others has now deciphered the process in some detail [53,54]. Using exogenous siRNAs they showed that the gene encoding the elongation factor 1α (EF1A) could be transcriptionally silenced. Transcriptional repression was associated with methylation of the EF1A promoter and required targeting of the siRNAs to the nucleus [54]. In addition, AGO1 (Argonaute 1), TRBP2 (TAR RNA-binding protein-2) and Polycomb protein EZH2 were shown to mediate transcriptional silencing of CCR5 (human immunodeficiency virus-1 co receptor) and tumour suppressor RASSF1A [55]. Interestingly, comparable studies revealed the unexpected finding that short RNAs could not only silence genes, but also activate their expression [56,57]. Activation was associated with increased di- and tri-methylation of histone H3K4 and reduced acetylation at histones H3K9 and H3K14. When the targets for several activating siRNAs were examined (progesterone receptor gene, E-cadherin, p21) it emerged that all genes were bi-directionally transcribed and the activating oligonucleotides targeted in fact the respective antisense transcripts [49,58].

Figure 1 | Endo-siRNAs from sense/antisense transcripts in mouse and zebrafish

The sense transcript encodes a sodium/phosphate transporter, the antisense transcript is most likely non-protein-coding. Northern blot analysis of RNA extracted from different mouse tissues (left-hand panel) and zebrafish embryos (right-hand panel) shows the expression of endo-siRNAs. The orientation of the oligonucleotides is tissue- or development-specific. The molecular mass in kDa is indicated. M, molecular mass marker lane.
Indeed, suppression of the p21 antisense transcript resulted in a loss of suppressive histone modifications in the sense promoter [49]. These findings were recently shown to apply to other mammals, such as monkeys and rodents [59]. A strong link between NATs and transcriptional gene silencing is suggested by the correlation of NATs and genes that show random monoallelic expression [38].

This, to a certain extent, closes a circle and explains the initial observation that NATs are under-represented on the X chromosome. Because only one copy of the X chromosome is usually active, NAT-induced silencing would trigger complete gene knockdown with potentially fatal consequences for the affected cell. If a comparable incident happened on autosomes, the second allele would compensate for the loss of function. This argument is supported by the fact that NATs expressed from mammalian X chromosomes tend to be clustered in areas that escape inactivation [29].

What good are NATs for? A hypothesis

To conclude this review, we attempt to answer the initial question and suggest a hypothesis of a higher-order biological role for NATs (Figure 2). The model predicts that sense and antisense transcripts are transiently co-expressed. The transcripts either persist as mRNAs or, alternatively, hybridize and initiate the processing into endo-siRNAs. Based on indirect evidence we hypothesize that the selection of a specific endo-siRNA strand is related to the formation of the RISC (RNA-induced silencing complex) with a non-slicing Argonaute protein [60]. Interestingly, the expression of the non-slicing Ago4 protein in testis mirrors the pattern established for NATs [61]. Recent investigations into RISC assembly revealed that the unwinding of the siRNA duplex after Ago1 binding is thermodynamically controlled [60]. We therefore hypothesize that double-stranded endo-siRNAs bind in random orientation to Ago4 to form a ‘pre-RISC’. If the fully processed mRNA is accessible in the cytoplasm, a RISC will be formed and contained in cytoplasmic compartments. If the target is unavailable in the cytoplasm, the pre-RISC may diffuse into the nucleus, target the nascent complementary transcript and induce a silencing response. It is conceivable that the RISC complexes in cytoplasm and nucleus have different stabilities and fates resulting in the detection of skewed single-stranded endo-siRNA ratios. To summarize, strand selection will be determined by the availability of stable mRNAs complementary to the endo-siRNAs in the cytoplasm. Any mutation that reduces cytoplasmic mRNA levels by, for example, affecting stability or nuclear export may induce selection of the opposite endo-siRNA strand and re-direction of transcriptional silencing.

The impact of the suggested model becomes tangible when the hypothesis outlined in the previous paragraph is considered in conjunction with the expression pattern of NATs. NATs are most prominently expressed in testis...
and, at low levels, in other tissues. In haploid cells in testis (spermatids), mutations that trigger NAT-induced gene silencing will cause a complete gene knockdown. As a consequence, mutated (‘unfit’) cells will be preferentially eliminated and a positively selected sperm population will emerge. In diploid cells, on the other hand, the mutagenic environment will induce progressive monoallelic gene expression. The random but cumulative nature of these changes in slowly regenerating somatic cells may contribute to senescence.

Essential questions remain to be answered; for example, how allele specificity of the entire process is achieved. In addition, elements of the model that are based on indirect evidence require robust experimental scrutiny. Nevertheless, endo-siRNAs and transcriptional gene silencing seem to make sense in the context of antisense.

Funding

This work is supported by the Dunhill Medical Trust.

References

7 Ding, S.W. and Voinnet, O. (2007) Antiviral immunity directed by small RNAs. Cell 130, 413-426
17 Ding, S.W. and Voinnet, O. (2007) Antiviral immunity directed by small RNAs. Cell 130, 413-426

© The Authors Journal compilation ©2010 Biochemical Society

Received 19 February 2010
doi:10.1042/BST0381144