Review TRENDS in Biochemical Sciences Vol.29 No.2 February 2004

In search of antisense

Giovanni Lavorgna[1], Dvir Dahary[2], Ben Lehner[3], Rotem Sorek[2], Christopher M. Sanderson[3] and Giorgio Casari[1]

[1]Human Molecular Genetics Unit, Dibit-San Raffaele Scientific Institute, Via Olgettina 58, 20132 Milan, Italy
[2]Compugen Ltd., 72 Pinchas Rosen St., Tel Aviv 69512, Israel
[3]MRC Rosalind Franklin Centre for Genomics Research, Hinxton, Cambridge, CB10 1SB, UK

In recent years, natural antisense transcripts (NATs) have been implicated in many aspects of eukaryotic gene expression including genomic imprinting, RNA interference, translational regulation, alternative splicing, X-inactivation and RNA editing. Moreover, there is growing evidence to suggest that antisense transcription might have a key role in a range of human diseases. Consequently, there have been several recent attempts to identify novel NATs. To date, ~2500 mammalian NATs have been found, indicating that antisense transcription might be a common mechanism of regulating gene expression in human cells. There are increasingly diverse ways in which antisense transcription can regulate gene expression and evidence for the involvement of NATs in human disease is emerging. A range of bioinformatic resources could be used to assist future antisense research.

Computational analysis of data from large-scale sequencing projects has revealed a surprising abundance of antisense transcripts in several eukaryotic genomes [1–6]. As some antisense transcripts have been shown to regulate gene expression [7,8], it is possible that antisense transcription might be a common mechanism of regulating gene expression in eukaryotic cells.

Natural antisense transcripts (NATs) are simply RNAs containing sequences that are complementary to other endogenous RNAs. They can be transcribed in *cis* from opposing DNA strands at the same genomic locus (*CIS*-NATS), or in *trans* from separate loci (*TRANS*-NATS). Because much of the work in the past few years has focussed on *cis*-NATs, these will be the subject of discussion in this review. It is important, however, to remember that *trans*-NATs can induce gene silencing in *Drosophila* [9] and that they might also function in humans [10]. Two other classes of *trans*-acting noncoding RNA are related to *trans*-NATs because they recognize their target RNAs by imprecise base-pairing: MICRO RNAS (miRNAs; see Glossary), which inhibit the translation of mRNAs [11], and SMALL NUCLEOLAR RNAS (snoRNAs), which guide the modification of noncoding RNAs [12]. Intriguingly, many vertebrate mRNAs contain highly conserved repetitive sequences or transposable elements that might also form *trans*-NATs [1].

Before the sequence of the human genome became available, a few human NATs had been identified by groups studying specific genomic loci. As a result of these observations, the authors of two excellent early reviews speculated that antisense transcripts might have an important and diverse role in the regulation of human gene expression [7,8]. However, the first evidence that antisense transcription was a common feature of eukaryotic genomes came from the analysis of reverse complementarity between all available human mRNA sequences [1]. This study identified 87 genomic loci encoding natural antisense transcripts, and predicted that >800 coding antisense transcripts would exist in the human genome. Subsequent studies both confirmed and extended these observations using databases of mRNAs [2] and expressed sequence tags (ESTs) [3]. In particular, Yelin *et al.* [5] have identified 2667 human NATs of which >1600 are predicted to be true NATs. More recently, analysis of many fully

Glossary

***Cis* and *trans*-NATs:** natural antisense transcripts (NATs) are RNAs that contain sequence that is complementary to other endogenous RNAs. They might be transcribed in *cis* from opposing DNA strands at the same genomic locus or in *trans* from separate loci.

double-stranded RNAs (dsRNAs): this variable-size RNA form, deriving from the interaction of antisense partners, can affect gene expression via RNA editing or RNA interference mechanisms.

microRNAs (miRNAs): these very small, *trans*-acting RNA molecules are processed, like small interfering RNAs, from larger dsRNA-precursor molecules. They control gene expression by binding to complementary sites in target mRNAs, thereby, regulating their stability or translation.

small interfering RNAs (siRNAs): short RNAs of ~22 nucleotides that mediate RNA interference. Like miRNAs, they are processed by larger dsRNAs. However, there are significant differences between these molecules, including the fact that miRNAs regulate the expression of genes at another locus, whereas siRNAs regulate the locus from which their sequence derives.

small nucleolar RNAs (snoRNAs): *trans*-acting RNA molecules that guide the modification of noncoding RNAs. They represent an abundant, evolutionarily ancient group of noncoding RNAs that possess impressively diverse functions ranging from 2′-*O*-methylation and pseudouridylation of various classes of RNAs, through nucleolytic processing of rRNAs to the synthesis of telomeric DNA.

RNA editing: post-transcriptional change of a gene-encoded sequence at the RNA level, excluding alterations owing to processes such as mRNA splicing and polyadenylation. It can occur either by the insertion or deletion of nucleotides or by the substitution of bases by modification.

RNA interference (RNAi): a recent technological advance that enables researchers to reduce gene expression at the post-transcriptional level. This form of RNA silencing is initiated by dsRNA expressed in or introduced into a cell of interest, which, upon cleavage in 21–23-nucleotide duplexes (termed siRNAs) by the enzyme Dicer, triggers homology-dependent degradation of the corresponding mRNA.

RNA masking: formation of RNA duplexes between sense and antisense transcripts that cause the masking of key regulatory features within either transcript, thereby, inhibiting the binding of important *trans*-acting factors. This can affect any step in gene expression involving protein–RNA interactions, including mRNA splicing, transport, polyadenylation, translation and degradation.

Corresponding author: Giovanni Lavorgna (giovanni.lavorgna@hsr.it).

 0968-0004/$ - see front matter doi:10.1016/j.tibs.2003.12.002

sequenced mouse cDNAs has predicted the existence of as many as 2500 mammalian *cis*-NATs [6]. Moreover, 1027 NATs have been identified in the *Drosophila*genome [4] and a similar number are predicted to exist in the *Caenorhabditis elegans* genome (K. Numata and M. Tomita, pers. commun.).

Functions of antisense transcripts

Although recent studies have identified many NATs, our understanding of how antisense transcription regulates gene expression in human cells remains largely incomplete. However, pioneering studies in several eukaryotic systems have identified three general mechanisms by which antisense transcription can regulate gene expression: transcriptional interference, RNA MASKING and DOUBLE-STRANDED RNA (dsRNA)-dependent mechanisms (Figure 1).

Transcriptional interference

Transcription by RNA polymerase II involves both large protein complexes and the unwinding of duplex DNA. As such, it is unlikely that two overlapping transcriptional units could be transcribed concomitantly. The effects of transcriptional interference at antisense loci were investigated in *Saccharomyces cerevisiae* by rearranging the *GAL10* and *GAL7* genes so that they are transcribed convergently [13]. When arranged convergently, but not overlapping, both genes are transcribed at normal levels. However, when the two transcriptional units overlap, steady-state mRNA levels are severely reduced due to an inhibition in transcription elongation [13]. These data suggest that the expression of *cis*-NAT partners could be tightly regulated through a process of competitive transcriptional interference (Figure 1a). Under these conditions *cis*-NATs might be expected to exhibit reciprocal

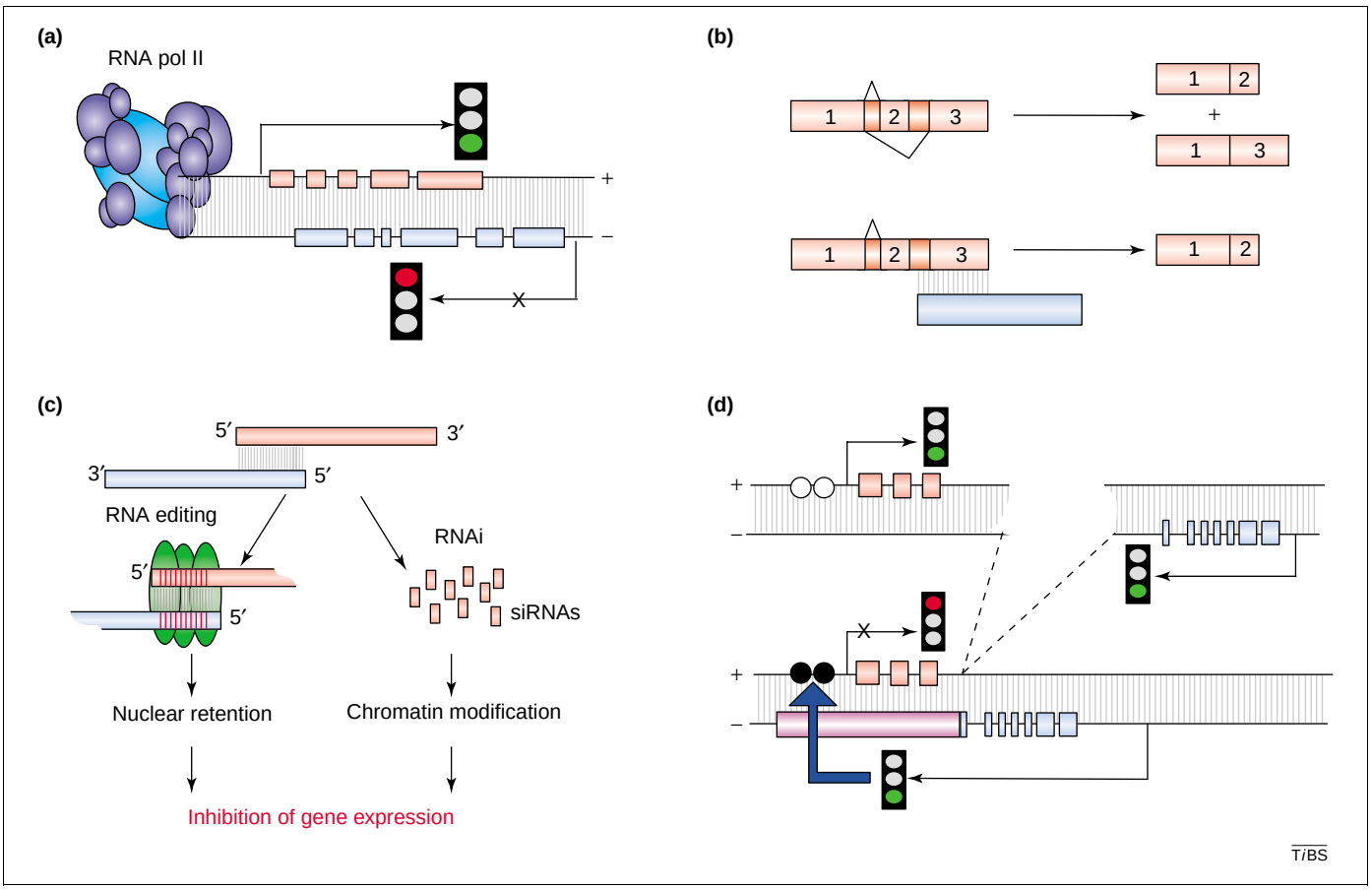

Figure 1. Different ways in which natural antisense transcripts (NATs) are known to regulate gene expression in eukaryotic cells. **(a)** Transcriptional interference: transcription of a gene encoded on the (+) strand of the DNA inhibits concomitant transcription of overlapping genes encoded on the (−) strand. This is because this would involve the simultaneous presence of two voluminous RNA pol II complexes on opposite strands, leading (as transcription proceeds) to the phenomena of collision and/or stalling, which would ultimately affect the activity of one or both protein complexes. **(b)** RNA masking: interaction of sense (red) and antisense transcripts (blue) might mask regulatory elements within the pre-mRNA that are required for exon selection and intron (dark red) removal. Consequently, expression of antisense transcripts can dictate the way in which the sense transcript is differentially spliced. **(c)** Double-stranded RNA (dsRNA)-dependent mechanisms: interaction between sense (red) and antisense transcripts (blue) results in the formation of dsRNA. Human cells have evolved mechanisms of avoiding the presence of dsRNAs in the cytoplasm of cells. In cases where the length of complementarity between sense and antisense transcripts is large (>200 nucleotides) double-stranded regions become extensively modified (red lines) by the RNA-editing machinery. Hyper-edited transcripts are then recognized by proteins (green), which inhibit their export from the nucleus and prevent their translation in the cytoplasm. Alternatively, dsRNAs might be digested into small fragments [small interfering RNAs (siRNAs)] by the RNA interference (RNAi) machinery. This process not only results in the destruction of the sense transcript, but the resulting fragments might also target chromatin remodelling factors to the genomic locus that encodes the sense transcript. The resulting changes in chromosome structure inhibit transcription at that locus. **(d)** Antisense-induced methylation: as a result of a chromosomal deletion the antisense transcript (blue) is brought in close proximity to the sense transcript (red). In addition, the antisense transcript becomes extended (magenta) so that it overlaps the sense transcript on the (+) strand. By some unknown mechanism, expression of the antisense transcript induces the methylation (black circles) of methylation-sensitive sites (white circles) in the CpG island upstream of the sense gene, thereby, inhibiting its transcription.

expression, which is true for antisense partners at the eukaryotic initiation factor 2α (eIF2α) [14], insulin-like growth factor type-2 receptor/antisense Igf2r RNA (Igf2r/Air) [15] and α1 (I) collagen [16] loci. For example, the gene encoding eIF2α has an associated antisense RNA transcribed from an initiator (Inr) element within the first intron, and transcription of the sense and antisense transcripts are inversely correlated at different stages of the cell cycle [17]. Moreover, deletion or mutation of the Inr element from reporter constructs decreases antisense transcription and increases sense transcript expression [14]. However, at other loci, sense and antisense transcription is not observed to be mutually exclusive, suggesting that alternative regulatory mechanisms must also exist [7,8]. For example, in myelin-deficient mice (*mld*) an abnormal antisense RNA drastically reduces expression of myelin basic protein (MBP) without any significant effect on MBP transcription [18]. In this case, it is more probable that regulation of gene expression might result from the direct interaction of antisense transcripts.

RNA masking

Formation of RNA duplexes between sense and antisense transcripts might mask key regulatory features within either transcript, thereby inhibiting the binding of important *trans*-acting factors (Figure 1b). This form of steric inhibition could affect any step in gene expression involving protein–RNA interactions, including mRNA splicing, transport, polyadenylation, translation and degradation. An example of this method of antisense regulation is the inhibition of alternative splicing induced by the Rev-ErbAα transcript, which overlaps one of two functionally antagonistic splice forms of the thyroid hormone receptor ErbAα mRNA (ErbAα2 not ErbAα1). In different B-cell lines, expression of Rev-ErbAα correlates with an increase in the ratio of ErbAα1- to ErbAα2-mRNA levels [19]. Moreover, expression of Rev-ErbAα inhibits production of ErbAα2 from an ErbAα minigene in cells [20], and antisense RNAs complementary to the ErbAα2-specific exon efficiently and specifically block ErbAα2 splicing *in vitro* [21]. These results show that an antisense RNA can specifically inhibit the alternative splicing of an mRNA, probably by blocking the accessibility of *cis* regulatory elements in the RNA.

dsRNA-dependent mechanisms and RNA interference

In addition to RNA masking, there is now strong evidence that the interaction of antisense partners can also affect gene expression via the activation of dsRNA-dependent pathways (Figure 1c). These might include RNA EDITING, or RNA INTERFERENCE (RNAi)-dependent gene silencing. In the first scenario, nuclear adenosine deaminases (ADARs) bind dsRNA and catalyze the hydrolytic deamination of adenosines to inosines [22]. Minimal editing of short (<15 bp) stem–loop duplexes enables the coding potential of an mRNA to be changed. However, long (>100 bp) perfect duplexes, similar to those that could result from antisense transcription, are hyper-edited so that ~50% of the adenosines on each strand are deaminated [22]. As a result, these RNAs might be either retained in the nucleus by the 54-kDa nuclear RNA-binding protein (p54nrb) multiprotein complex [23] or degraded by a cytoplasmic endonuclease [24]. Long RNA duplexes from the mouse polyoma virus [25] and the *Drosophila 4f-rnp* gene [26] are hyper-edited and retained in the nucleus. However, this form of editing has not yet been demonstrated for any mammalian dsRNAs, so the generality of these mechanisms remains to be established.

The formation of dsRNAs might also induce gene silencing via RNAi pathways. When dsRNA is introduced into most eukaryotic cells it is efficiently cleaved by the enzyme Dicer into 21–23 nucleotide duplexes, termed SMALL INTERFERING RNAS (siRNAs). These fragments then target the specific destruction of homologous mRNAs [27]. There are two precedents to suggest that sense–antisense transcription can induce gene silencing via an RNAi-dependent mechanism: (i) silencing of the *Drosophila* Stellate repeats [9] and (ii) the formation of heterochromatin in *Schizosaccharomyces pombe* [28]. Prevention of hyper-expression of the *Drosophila* Stellate repeats in testis is essential for male fertility and requires the homologous Su(Ste) tandem repeats. Aravin *et al.* [9] found that the Su(Ste) repeats produce both sense and antisense RNAs that form dsRNA *in vivo* and are cleaved to form heterogeneous 25–27-nt RNA species. Significantly, transfected *Su(Ste)* dsRNA was able to eliminate *Stellate* transcripts in cell culture and also reduced levels of the *Su(Ste)* sense RNA via a negative feedback loop. *Stellate* silencing is dependent on the two proteins – aubergine and spindle-E – which are necessary for RNAi in the germline [29]. Mutation in *spindle-E* also results in de-repression of other genomic tandem repeats and retrotransposons, suggesting that sense–antisense transcription and RNAi together might be a general mechanism of endogenous repetitive-element silencing in *Drosophila*. A similar mechanism is proposed to explain the silencing of heterochromatin in *S. pombe* and, in this situation, silencing has been shown to be at the level of transcription. Overlapping transcription from centromeric and interspersed repetitive elements produces dsRNA, which is cleaved by the RNAi machinery and guides recruitment of heterochromatin proteins to the repetitive elements and subsequent transcriptional silencing [28,30]. This mechanism has also been demonstrated to result in the silencing of single-copy meiotic-specific genes [30]. Whether endogenous vertebrate NATs can induce RNAi and gene silencing remains to be established, but such a mechanism could be used to either maintain low steady-state levels of transcription, or turn off production of a sense RNA by induction of an antisense transcript.

In addition to these examples, there are other instances in which antisense transcription is known to affect gene expression but the underlying mechanism of action remains unclear. This is particularly true for antisense transcripts involved in gene imprinting [15,31] or the methylation of CpG islands and chromatin remodelling [32].

Relevance of antisense transcription to disease

Given the diverse ways in which NATs can affect the expression of eukaryotic genes, it is hardly surprising that changes in antisense transcription can lead to abnormal

patterns of gene expression, which in turn contribute to pathological phenotypes. To date, there are only a few examples of antisense transcripts that are implicated in human disease. However, it is likely that these examples are only the tip of a rather large iceberg.

For several years it has been clear that many imprinted genes are associated with antisense transcripts. For example, research into the regulation of imprinted genes within the human *15q11–15q13* region have implicated the expression of a large SNURF-SNRPNsense/UBE3A (SNRPN upstream reading frame-small nuclear ribonucleoprotein N/E6-AP ubiquitin-protein ligase) antisense transcript with the reduction in UBE3A expression which is associated with Prader–Willi and Angelman syndromes [33]. In addition, reduced expression of the exon 1B isoform of the basic fibroblast growth factor (bFGF) antisense transcript (1B FGF-AS) is thought to contribute to the proliferation of endometrial cells observed in patients with endometriosis [34]. Another mechanism by which antisense transcription might contribute to disease is the generation of abnormal antisense transcripts that result from chromosomal rearrangements. In a recent report, Tufarelli *et al.* [32] describe a novel disease mechanism that leads to an inherited form of α-thalassemia via silencing of the hemoglobin α-2 (*HBA2*) gene by a *cis*-acting antisense RNA. This study shows that in the affected individual, as well as in a transgenic model and differentiating embryonic stem cells, transcription of the antisense RNA is associated with *de novo* methylation of the HBA2 CpG island (Figure 1d). However, despite a clear correlation between antisense transcription, CpG-island methylation and the localized modification of chromatin structure, the molecular mechanism underlying this method of gene silencing remains unclear. Nevertheless, it is intriguing to ask whether other examples of this mechanism could already exist. Tufarelli *et al.* also suggest that, in malignant cells (in which the bulk of the genome becomes hypomethylated), the resulting de-repression of short interspersed repeated sequences (SINEs), long interspersed repeated sequences (LINEs) and retroviral sequences [35,36] might generate transcriptional 'noise'. Thereby, aberrant antisense RNA transcripts are yielded, some of which might randomly initiate methylation of CpG islands associated with key oncogenes and tumor suppressors. This would then inhibit their expression and promote the evolution of malignant clones. Interestingly, data obtained by Shendure and Church [3] also corroborate the possible connection between the up-regulation of antisense transcription and cancer. By analyzing ESTs belonging to the human gene encoding cell-death inducing DFFA-like effector B (CIDEB) a potential tumor suppressor, they were able to compute the fraction of sense versus antisense ESTs that were derived from different tissues (http://arep.med.harvard.edu/twister/antisense/Hs.288835.html). Significantly, they found that neoplastic tissues express a greater fraction of antisense ESTs than corresponding sense ESTs (Figure 2). Because the sense transcript encodes a pro-apoptotic gene, this result suggests that an increase in antisense RNA transcription in cancer tissues might act to suppress the expression of potential tumor-suppressor genes. A similar pattern of

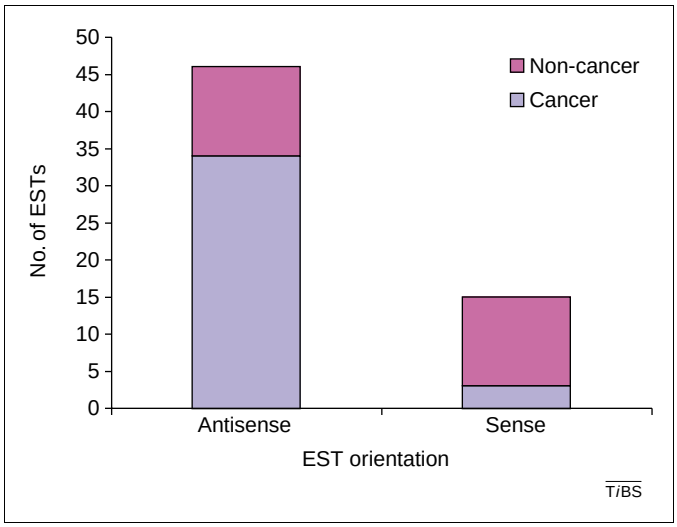

Figure 2. Possible relevance of antisense transcription to cancer, as revealed by computational means. Sense and antisense expressed sequence tags (ESTs) from Unigene cluster Hs.288835, containing the potential tumor-suppressor human CIDEB (cell-death-inducing DFFA-like effector B) gene, were plotted against orientation. The *y*-axis indicates the number of sense or antisense ESTs observed in the CIDEB cluster, and the relative proportions arising from neoplastic versus non-neoplastic tissues are indicated. A significantly greater fraction of the antisense ESTs (34/46 = 0.74) than the sense ESTs (3/15 = 0.3) were derived from neoplastic tissues ($p = 0.0002$ by χ-squared statistic), suggesting the hypothesis that upregulation of the antisense RNA species in cancer tissues has functional relevance with respect to suppression of this sense transcript. Reproduced, with permission, from Ref. [3].

sense and/or antisense EST expression was also observed for the gene encoding Burkitt lymphoma receptor 1 (http://arep.med.harvard.edu/twister/as_main.htm#hs113916).

Although more examples are needed to fully understand the role of antisense transcription in cancer, the available evidence suggests that a systematic genome-wide *in-silico* search for antisense transcripts that are enriched in neoplastic tissues could facilitate the identification of new tumor-suppressor genes. The AntiHunter software tool (http://bio.ifom-firc.it/ANTIHUNTER/) appears to be well suited for this task because it allows the source and neoplastic state of tissues to be identified for each EST. In addition, the AntiHunter tool can be set to search for complementary antisense ESTs, thereby, enabling the relative expression of both partners to be assessed under different conditions in different tissues.

Web resources for *in-silico* detection of antisense transcription

The availability of the complete human genome sequence and the accumulation of millions of expressed sequences (mRNAs and ESTs) have made it possible for large-scale predictions of naturally occurring antisense transcription. Indeed, in the past two years, several studies have used these sources of information in attempts to produce comprehensive datasets of sense–antisense pairs. The results of these studies are summarized in Table 1.

The first study, conducted by Lehner *et al.* [1], took the approach of direct BLAST comparisons between the ~12 000 human full-length mRNAs that were available at the time, searching for regions of complementarity between pairs of mRNAs. By mapping mRNAs with such complementarity to the human genome, Lehner *et al.*

Table 1. Web resources for *in-silico* detection of antisense transcription[a]

Tool	Input	Predicted antisenses available	Organism	Browser included	URL	Refs
Db	Human mRNAs (RefSeq and full CDS)	87 human	Human	No	www.hgmp.mrc.ac.uk/Research/Antisense	[1]
Db	UniGene build 146 (human) and build 100 (mouse)	144 human, 73 mouse	Human and mouse	Yes	http://arep.med.harvard.edu/twister/antisense.html	[3]
Db	mRNAs and ESTs from gb125	~2600	Human	No	www.labonweb.com/antisense	[5]
Db	RIKEN full-length mRNAs (FANTOM2)	~2500	Mouse	Yes	http://genome.gsc.riken.go.jp/m/antisense/	[6]
Sw	Current dbEST	Dynamic, depends on user query and EST availability	All organisms for which there are ESTs	No	http://bio.ifom-firc.it/ANTIHUNTER/	

[a]Abbreviations: CDS, protein-coding sequence; db, database; EST, expressed sequence tag; FANTOM2.

identified 87 sense–antisense pairs (appears as a searchable table with links to graphical representations of sense–antisense regions on their website). This set of antisense cases is considered reliable because it is based on full-length mRNA sequences. However, it represents a very small set of the actual number of human antisenses because of the limited quantity of input sequences analyzed.

Two other studies attempted to identify antisense transcription using ESTs as input in addition to mRNAs. On one hand, ESTs add two orders of magnitude of input information – although there are currently more than five million human ESTs in GenBank, only a few tens of thousands of full-length mRNAs exist there. In addition, mammalian genes frequently contain alternative 3′ ends [37] and alternative promoters [38], which are underrepresented in the dataset of full-length mRNAs and are, therefore, mostly detectable through analysis of ESTs. On the other hand, a significant portion of dbEST (database of ESTs) is thought to represent artifacts rather than true expressed sequences [39]. Moreover, many EST libraries are non-directional, and therefore the orientation of an EST cannot be easily determined. Furthermore, in some cases, the orientation of ESTs is mis-annotated even when a direction annotation exists. For these reasons, antisense prediction using ESTs requires more sophisticated tools.

Shendure and Church [3] were the first to use ESTs for antisense detection, finding 144 human and 73 mouse antisense cases. To avoid the inherent problems associated with EST data and to reduce the rate of false positives, they employed restrictive parameters when selecting the ESTs for analysis. They first selected only a subset of EST libraries (~900 human and ~200 mouse libraries), for which they found high reliability of EST direction annotation (for comparison, dbEST currently contains >6500 human libraries [3]). To further improve the prediction, Shendure and Church analyzed only genomic loci with at least one full-length mRNA for which the correct transcriptional annotation was known. Then, loci in which there was a significantly high frequency of ESTs oppositely oriented to the full-length mRNA were identified, and the sequences in these loci were mapped to the genome. The final set of antisense cases comprised loci that appeared to contain two distinct but oppositely oriented RNA species, according to exon–intron structures, poly(A) signals and tail locations.

The antisense genes found by Shendure and Church are summarized in a table at their website (http://arep.med.harvard.edu/twister/antisense.html), and each case can be viewed graphically. The graphic viewer presents all the expressed sequences aligned to the genomic locus, grouped according to their orientation. The viewer also enables visualization of human-mouse genomic sequence identity levels at antisense loci, allowing assessment of the functional importance of the overlapping regions. Apart from the FANTOM2 mouse antisense analysis (http://fantom2.gsc.riken.go.jp), this is the only website providing a graphical access to curated antisense data.

Yelin *et al.* [5] also used ESTs for antisense prediction with their 'Antisensor' algorithm but, unlike Shendure and Church, they used the entire set of publicly available ESTs as input. To determine the orientation of individual sequences, Antisensor used splicing patterns, poly(A)-tail orientations, protein-coding sequence (CDS) annotation and other parameters. Orientation was assigned to sequences for which one or more of these parameters existed. Finally, loci in which there were overlapping sequences with opposite inferred orientations were considered antisense cases. The final output of the Antisensor was a set of ~2600 gene pairs presented as a searchable table at the Antisensor website (www.labonweb.com/antisense). In addition, this set can be queried using a web interface that allows the user to find whether a sequence (EST or mRNA) is mapped to one of the identified antisense loci.

Although the methods used by Yelin *et al.* resemble those used by Shendure and Church, there are several key differences between these two analyses. First, as mentioned above, Shendure and Church employed more constraints on their initial input, and used only ~10% of the available EST libraries. In addition, whereas Shendure and Church analyzed only genomic loci in which there was at least one mRNA, Yelin *et al.* also used genomic loci in which no full-length mRNA was present. As a result, the dataset of Yelin *et al.* is one order of magnitude larger than that of Shendure and Church. However, the rate of

false positives in the Yelin data are estimated to be higher: in an experimental validation, 85% of the antisense cases of Shendure and Church were validated, compared with only 60% of the Yelin set.

The recent expansion in mouse full-length cDNA data from the FANTOM2 project [40] enabled Kiyosawa *et al.* [6] to compile a comprehensive dataset of antisense cases without using ESTs. As input, this group used a set of ~61 000 full-length FANTOM2 cDNA sequences in addition to the full-length mouse mRNA that already existed in GenBank. By mapping these cDNAs to the draft mouse genome, Kiyosawa *et al.* were able to find ~2500 cDNA pairs originating from opposite strands of the same genomic loci. Although this analysis resembles the analysis performed by Lehner *et al.* on human cDNAs, it yielded more predicted antisense cases. Two main reasons probably contributed to this difference: (i) the larger input used by Kiyosawa *et al.* and (ii) the FANTOM2 cDNA dataset contains many more non-protein-coding RNAs than the RefSeq database. More than 70% of the antisense pairs identified by Kiyosawa *et al.* involve a non-coding RNA as one of the members in the pair.

The dataset of 2500 antisense pairs obtained by Kiyosawa is accessible either by scanning a genomic region, or through an accession number or the gene description. The sense–antisense pairs are presented in a viewer showing the exon–intron structure of each of the members in the pair, as well as the GC-content and putative promoters at the genomic locus.

One limitation of all these databases is that they are the output of an analysis carried out at a specific point in time. The EST databases, as well as the databases of full-length cDNAs, are constantly growing and have the potential to add essential new data for detection of antisense transcription. This new data will, of course, be missing from the above-described antisense databases. To address this, Lavorgna *et al.* developed an online antisense detection tool, 'AntiHunter' [41]. Given a genomic sequence with an annotation of its expressed regions, this tool finds ESTs that are in the antisense orientation to the annotated input sequence. As sources of orientation information, the AntiHunter software uses the database annotation of the sequence, as well as the splice junctions and the presence of a poly(A) tail in ESTs. In addition, AntiHunter can analyze genomic regions from any species for which there are ESTs and genomic data available, whereas the other web-based resources are restricted to specific organisms. However, AntiHunter is a software tool, as opposed to a database tool, and, therefore, its work cannot be pre-calculated. Consequently, response times are significantly longer, for example, to perform a query using a 500-Kb sequence might take a few hours, whereas the answer from the other tools is almost instantaneous.

All of the web-based resources described here are specifically aimed at finding natural antisense transcripts. However, additional general resources that are not specially designed for this purpose can also help in antisense transcription detection. Among these resources are the popular online genome browsers, such as the University of California Santa Cruz (UCSC; http://genome.ucsc.edu/) and Ensembl (http://www.ensembl.org/) browsers, in which the user can view a specific genomic locus with all cDNAs and ESTs aligned to it. The direction of full-length cDNAs and spliced ESTs is usually shown so that the user can determine whether there is an antisense overlap with a chosen gene. These browsers are not perfect, however, because they do not present several key orientation parameters such as poly(A) sequences or sites, and database annotation. Thus, the UCSC and Ensembl genome browsers can be mainly used to view the results of the antisense detection tools in their genomic context. In this sense, they are an important complement of all the tools discussed above.

Concluding remarks

A major challenge for the future will be to establish which of the identified human NATs actually affect gene expression and by which mechanism. Ideally, these questions should be addressed in parallel for many NATs. The mechanisms discussed in this review make predictions that are amenable to large-scale parallel analysis. For example, RNAi mechanisms would produce endogenous 21–23-nt RNAs derived from the regions of overlapping transcription, and these could be identified using cloning strategies similar to those used to identify miRNAs [42]. Equally, if transcriptional interference is a widespread mechanism, then sense and antisense transcripts should have reciprocal cellular expression patterns. Theoretically, this could be investigated by mining available microarray data, although as coverage of antisense pairs is limited in many public microarray datasets, informative large-scale studies might require the use of custom-designed microarrays.

Although the search for NATs is still in its infancy, *in-silico* methods for detecting potential antisense mRNA transcripts have proved to be an effective approach for the identification of new members of this class of transcripts. Surprisingly, even error-prone material, such as EST sequences, have been shown to be a reliable (and frequently updated) source of potential antisense transcripts once robust algorithms that filter out noisy data have been designed.

Significantly, the availability of many examples of human NATs and the development of tools to search for antisense transcripts that are associated with disease loci should provide a better insight into the functions of different antisense transcripts and the molecular pathology associated with many important human diseases.

Acknowledgements

We thank Gisela Storz, Doug Higgs and Galit Rotman for helpful comments and suggestions.

References

1 Lehner, B. *et al.* (2002) Antisense transcripts in the human genome. *Trends Genet.* 18, 63–65
2 Fahey, M.E. *et al.* (2002) Overlapping antisense transcription in the human genome. *Comp. Functional Genomics* 3, 244–253
3 Shendure, J. and Church, G.M. (2002) Computational discovery of sense–antisense transcription in the human and mouse genomes. *Genome Biol.* 3, 1–14
4 Misra, S. *et al.* (2002) Annotation of the *Drosophila melanogaster* euchromatic genome: a systematic review. *Genome Biol.* 3, 83

5 Yelin, R. *et al.* (2003) Widespread occurrence of antisense transcription in the human genome. *Nat. Biotechnol.* 21, 379–386
6 Kiyosawa, H. *et al.* (2003) Antisense transcripts with FANTOM2 clone set and their implications for gene regulation. *Genome Res.* 13, 1324–1334
7 Vanhee-Brossollet, C. and Vaquero, C. (1998) Do natural antisense transcripts make sense in eukaryotes? *Gene* 211, 1–9
8 Kumar, M. and Carmichael, G.G. (1998) Antisense RNA: function and fate of duplex RNA in cells of higher eukaryotes. *Microbiol. Mol. Biol. Rev.* 62, 1415–1434
9 Aravin, A.A. *et al.* (2001) Double-stranded RNA-mediated silencing of genomic tandem repeats and transposable elements in the *D. melanogaster* germline. *Curr. Biol.* 11, 1017–1027
10 Morelli, S. *et al.* (1997) The antisense bcl-2-IgH transcript is an optimal target for synthetic oligonucleotides. *Proc. Natl. Acad. Sci. U. S. A.* 94, 8150–8155
11 Brennecke, J. and Cohen, S.M. (2003) Towards a complete description of the microRNA complement of animal genomes. *Genome Biol.* 4, 228
12 Kiss, T. (2002) Small nucleolar RNAs: an abundant group of noncoding RNAs with diverse cellular functions. *Cell* 109, 145–148
13 Prescott, E.M. and Proudfoot, N.J. (2002) Transcriptional collision between convergent genes in budding yeast. *Proc. Natl. Acad. Sci. U. S. A.* 99, 8796–8801
14 Silverman, T.A. *et al.* (1992) Role of sequences within the first intron in the regulation of expression of eukaryotic initiation factor 2α. *J. Biol. Chem.* 267, 9738–9742
15 Wutz, A. *et al.* (1997) Imprinted expression of the *Igf2r* gene depends on an intronic CpG island. *Nature* 389, 745–749
16 Farrell, C.M. and Lukens, L.N. (1995) Naturally occurring antisense transcripts are present in chick embryo chondrocytes simultaneously with the down-regulation of the α 1 (I) collagen gene. *J. Biol. Chem.* 270, 3400–3408
17 Noguchi, M. *et al.* (1994) Characterization of an antisense Inr element in the eIF-2α gene. *J. Biol. Chem.* 269, 29161–29167
18 Tosic, M. *et al.* (1990) Post-transcriptional events are responsible for low expression of myelin basic protein in myelin deficient mice: role of natural antisense RNA. *EMBO J.* 9, 401–406
19 Hastings, M.L. *et al.* (1997) Expression of the thyroid hormone receptor gene, erbAα, in B lymphocytes: alternative mRNA processing is independent of differentiation but correlates with antisense RNA levels. *Nucleic Acids Res.* 25, 4296–4300
20 Hastings, M.L. *et al.* (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
21 Munroe, S.H. and Lazar, M.A. (1991) Inhibition of c-erbA mRNA splicing by a naturally occurring antisense RNA. *J. Biol. Chem.* 266, 22083–22086
22 Bass, B.L. (2002) RNA editing by adenosine deaminases that act on RNA. *Annu. Rev. Biochem.* 71, 817–846
23 Zhang, Z. and Carmichael, G.G. (2001) The fate of dsRNA in the nucleus: a p54(nrb)-containing complex mediates the nuclear retention of promiscuously A-to-I edited RNAs. *Cell* 106, 465–475
24 Scadden, A.D. and Smith, C.W. (2001) Specific cleavage of hyper-edited dsRNAs. *EMBO J.* 20, 4243–4252
25 Kumar, M. and Carmichael, G.G. (1997) Nuclear antisense RNA induces extensive adenosine modifications and nuclear retention of target transcripts. *Proc. Natl. Acad. Sci. U. S. A.* 94, 3542–3547
26 Peters, N.T. *et al.* (2003) RNA editing and regulation of *Drosophila* 4f-rnp expression by sas-10 antisense readthrough mRNA transcripts. *RNA* 9, 698–710
27 Hannon, G.J. (2002) RNA interference. *Nature* 11, 244–251
28 Volpe, T.A. *et al.* (2002) Regulation of heterochromatic silencing and histone H3 lysine-9 methylation by RNAi. *Science* 297, 1833–1837
29 Kennerdell, J.R. *et al.* (2002) RNAi is activated during *Drosophila* oocyte maturation in a manner dependent on aubergine and spindle-E. *Genes Dev.* 16, 1884–1889
30 Schramke, V. and Allshire, R. (2003) Hairpin RNAs and retrotransposon LTRs effect RNAi and chromatin-based gene silencing. *Science* 301, 1069–1074
31 Reik, W. and Walter, J. (2001) Genomic imprinting: parental influence on the genome. *Nat. Rev. Genet.* 2, 21–32
32 Tufarelli, C. *et al.* (2003) Transcription of antisense RNA leading to gene silencing and methylation as a novel cause of human genetic disease. *Nat. Genet.* 34, 157–165
33 Runte, M. *et al.* (2001) The IC-SNURF-SNRPN transcript serves as a host for multiple small nucleolar RNA species and as an antisense RNA for UBE3A. *Hum. Mol. Genet.* 10, 2687–2700
34 Mihalich, A. *et al.* (2003) Different basic fibroblast growth factor and fibroblast growth factor–antisense expression in eutopic endometrial stromal cells derived from women with and without endometriosis. *J. Clin. Endocrinol. Metab.* 88, 2853–2859
35 Ehrlich, M. (2000) DNA hypomethylation and cancer. In *DNA Alterations in Cancer* (Ehrlich, M., ed.), pp. 273–291, Eaton
36 Baylin, S.B. and Herman, J.G. (2000) Epigenetics and loss of gene function in cancer. In *DNA Alterations in Cancer* (Ehrlich, M., ed.), pp. 293–309, Eaton publishing, MA, USA
37 Iseli, C. *et al.* (2002) Long-range heterogeneity at the 3′ ends of human mRNAs. *Genome Res.* 12, 1068–1074
38 Zavolan, M. *et al.* (2003) Impact of alternative initiation, splicing, and termination on the diversity of the mRNA transcripts encoded by the mouse transcriptome. *Genome Res.* 13, 1290–1300
39 Sorek, R. and Safer, H.M. (2003) A novel algorithm for computational identification of contaminated EST libraries. *Nucleic Acids Res.* 31, 1067–1074
40 Okazaki, Y. *et al.* (2002) Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs. *Nature* 420, 563–573
41 Lavorgna, G. *et al*. AntiHunter: searching BLAST output for EST antisense transcripts. *Bioinformatics* (in press)
42 Lagos-Quintana, M. *et al.* (2003) New microRNAs from mouse and human. *RNA* 9, 175–179