The complexity of antisense transcription revealed by the study of developing male germ cells

Wai-Yee Chan a,b,⁎, Shao-Ming Wu a, Lisa Ruszczynski a, Evelyn Law a, Tin-Lap Lee a, Vanessa Baxendale a, Alan Lap-Yin Pang a, Owen M. Rennert a

a Laboratory of Clinical Genomics, National Institute of Child Health and Human Development, National Institutes of Health, Building 49, Room 2A08, 49 Convent Drive, MSC 4429, Bethesda, MD 20892-4429, USA
b Department of Pediatrics, Georgetown University, Washington, DC 20007, USA

Received 22 October 2005; accepted 13 December 2005
Available online 3 February 2006

Abstract

Computational analyses have identified the widespread occurrence of antisense transcripts in the human and the mouse genome. However, the structure and the origin of the majority of the antisense transcripts are unknown. The presence of antisense transcripts for 19 of 64 differentially expressed genes during mouse spermatogenesis was demonstrated with orientation-specific RT-PCR. These antisense transcripts were derived from a wide variety of origins, including processed sense transcripts, intronic and exonic sequences of a single gene or multiple genes, intergenic sequences, and pseudogenes. They underwent normal and alternative splicing, 5′ capping, and 3′ polyadenylation, similar to the sense transcripts. There were also antisense transcripts that were not capped and/or polyadenylated. The testicular levels of the sense transcripts were higher than those of the antisense transcripts in all cases, while the relative expression in nontesticular tissues was variable. Thus antisense transcripts have complex origins and structures and the sense and antisense transcripts can be regulated independently.

Keywords: Antisense transcription; Intron; Exon; Intergenic; Pseudogene; Mouse; Spermatogonia; Spermatocytes; Spermatids

Although antisense transcription has been known to occur in prokaryotes for many years, the widespread occurrence of antisense transcripts in humans and mice has only recently been documented. Computational analyses estimated 8 to 20% of human and mouse genes form sense–antisense transcript pairs [1–4]. A more recent study of 10 human chromosomes indicated about 61% of surveyed loci have antisense transcripts [5]. The significance of the majority of antisense transcripts is currently unknown, though a number of biological functions have been proposed [6]. A role for antisense transcription in disease processes has also been suggested [7–10]. Despite this apparent biological importance of the antisense transcripts, very little is known of the mechanisms by which they are generated. Structural information of antisense transcripts is also very limited.

Furthermore, the occurrence of antisense transcription in specific biological processes has rarely been studied. A number of biological events in spermatogenesis such as genomic imprinting, translation repression, and stage-specific alternative splicing [11] are frequently associated with antisense transcripts. A systematic search for antisense transcripts in spermatogenic cells has not been reported. Recent profiling of expressed genes in mouse germ cells at different stages of development, including type A spermatogonia, pachytene spermatocytes, and round spermatids, by serial analysis of gene expression (SAGE) [12] offered a unique opportunity to examine the occurrence of antisense transcription during this critical process in development. This study documents the occurrence of antisense transcripts in these three types of germ cells. Characterization of antisense transcripts cloned revealed the complex origins and structural features of these RNA molecules. Expression studies showed that the sense and the antisense transcripts can be regulated independently.
Results and discussion

Orientation-specific reverse transcription-polymerase chain reaction (RT-PCR)

Sixty-four genes that had been confirmed to be differentially expressed in mouse germ cells at different stages of spermatogenesis by microarray [13] and/or quantitative real-time RT-PCR (QPCR) and corroborated by SAGE [12] were selected for this investigation. The list of the genes is shown in Supplementary Table S1. Thirteen differential expression patterns could be distinguished (results not shown). The most common pattern of expression was 0 or very low in spermatogonia and equally high in spermatocytes and spermatids (18 genes). The second most common pattern of expression was 0 or very low in spermatogonia, increased in spermatocytes, and peaked in spermatids (17 genes). The copy number of the representative SAGE tag of the majority of the genes (32 genes) was either 0 (33 genes) or very low in spermatogonia. To identify the antisense transcripts, SAGE tags matching the UniGene cluster of the genes were identified and multiple SAGE tags matching the same UniGene cluster were aligned with all mRNAs deposited in that UniGene cluster. Orientation of the transcript represented by the SAGE tag was determined in reference to the known mRNA and confirmed by orientation-specific RT-PCR [14].

Following alignment of the matching SAGE tags with sequences in their respective UniGene clusters, 41 genes (64%) were shown to have SAGE tags matching the antisense strand (Supplementary Table S1). The presence of sense–antisense overlapping transcript was confirmed for 19 genes with orientation-specific RT-PCR using total RNA as the template (Table 1). The results of analysis of the products of orientation-specific RT-PCR using polyacrylamide gel electrophoresis (PAGE) are shown in Supplementary Fig. S1. Among the rest of the 41 genes, 10 genes were considered to be not confirmed of having antisense transcripts because a band was present in the no-primer control of the RT-PCR. The remaining 12 genes were not analyzed by orientation-specific RT-PCR because the copy number of the antisense SAGE tags was extremely low. Expression of the antisense transcript was confirmed by cloning and sequencing of the amplicons of the orientation-specific RT-PCR. This, 19 of 64 (~30%) differentially expressed genes had antisense transcripts.

Examination of the nucleotide sequence of the antisense amplicons showed that they could be divided into three main groups based on the comparison with their sense genes. Fig. 1 shows the three groups of antisense amplicons. In group 1 the antisense amplicons were 100% complementary to the sense transcripts. This group could be divided into two subgroups. In subgroup 1A, the amplicon was contained in a single exon. This subgroup included the antisense transcript of a disintegrin and metalloprotease domain 5 (Adam5); diazepam-binding inhibitor-like 5 (Dbil5); DnaJ (Hsp40) homolog, subfamily B, member 3 (DnajB3); four and a half LIM domains 4 (Fhl4); glucokinase activity-related sequence 2 (Gk-rs2); phosducin-like 2 (Pdcl2); peptidylprolyl isomerase C (PpIC); γ isoform of globular C 30% differen-

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Gene name</th>
<th>SAGE tag No. in Spga–Sp cy–Spdt (S/AS)</th>
<th>Mean Δ cycle No. ± SD testicular total RNA</th>
<th>Approximate S/AS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ch10</td>
<td>Heat shock 10-kDa protein 1 (chaperone 10)</td>
<td>106–2–2/6–0–0</td>
<td>3.85 ± 0.35</td>
<td>14</td>
</tr>
<tr>
<td>Calm2</td>
<td>Calmodulin 2</td>
<td>110–304–40/6–20–0</td>
<td>4.07 ± 0.44</td>
<td>17</td>
</tr>
<tr>
<td>PpIC</td>
<td>Peptidylprolyl isomerase C</td>
<td>14–1–0/7–20–11</td>
<td>5.02 ± 0.75</td>
<td>32</td>
</tr>
<tr>
<td>Pdcl2</td>
<td>Phosducin-like 2</td>
<td>0–82–91/0–5–7</td>
<td>6.52 ± 0.22</td>
<td>92</td>
</tr>
<tr>
<td>Sh3–Stam</td>
<td>Associated molecule with the SH3 domain of STAM</td>
<td>1–7–0–0–0–7–0</td>
<td>7.72 ± 1.00</td>
<td>211</td>
</tr>
<tr>
<td>Tsg1</td>
<td>Testis-specific gene 1</td>
<td>0–310–419/0–31–51</td>
<td>8.70 ± 0.15</td>
<td>416</td>
</tr>
<tr>
<td>Gk-rs2</td>
<td>Glucokinase activity, related sequence 2</td>
<td>0–54–41/0–6–5</td>
<td>9.00 ± 0.32</td>
<td>512</td>
</tr>
<tr>
<td>Uba52</td>
<td>Ubiquitin A-52 residue ribosomal protein fusion product 1</td>
<td>531–113–203/20–4–13</td>
<td>9.20 ± 0.15</td>
<td>588</td>
</tr>
<tr>
<td>Tect3</td>
<td>T-complex-associated testis expressed 3</td>
<td>4–471–360/6–20–0</td>
<td>9.88 ± 1.09</td>
<td>942</td>
</tr>
<tr>
<td>Prm1</td>
<td>Proteamine 1</td>
<td>0–41–177/0–2–26</td>
<td>9.88 ± 0.20</td>
<td>942</td>
</tr>
<tr>
<td>Dnajb3</td>
<td>Dnaj (Hsp40) homolog, subfamily B, member 3</td>
<td>0–63–84/0–4–12</td>
<td>10.50 ± 0.21</td>
<td>1448</td>
</tr>
<tr>
<td>Ppp1cc</td>
<td>Protein phosphatase 1, catalytic subunit, γ isoform</td>
<td>82–363–382/5–18–6</td>
<td>10.70 ± 0.34</td>
<td>1663</td>
</tr>
<tr>
<td>Prm2</td>
<td>Proteamine 2</td>
<td>0–103–147/0–1–9</td>
<td>12.50 ± 0.12</td>
<td>5792</td>
</tr>
<tr>
<td>Dbl5</td>
<td>Dazip3: binding inhibitor-like 5</td>
<td>2–64–119/0–1–4</td>
<td>14.50 ± 0.13</td>
<td>23,170</td>
</tr>
<tr>
<td>Ldh3C</td>
<td>Lactate dehydrogenase 3, C chain, sperm specific</td>
<td>0–282–243/0–4–2</td>
<td>16.10 ± 0.89</td>
<td>70,239</td>
</tr>
<tr>
<td>Adam5</td>
<td>A disintegrin and metalloprotease domain 5</td>
<td>3–54–92/0–12–4</td>
<td>16.81 ± 1.54</td>
<td>114,898</td>
</tr>
<tr>
<td>Sap17</td>
<td>Sperm autoantigenic protein 17</td>
<td>2–116–377/1–21–30</td>
<td>17.47 ± 0.45</td>
<td>181,549</td>
</tr>
<tr>
<td>Fhl4</td>
<td>Four and a half LIM domain 4</td>
<td>0–164–272/2–19–27</td>
<td>19.54 ± 1.72</td>
<td>762,300</td>
</tr>
</tbody>
</table>

The germ cell SAGE tag numbers are from Ref. [12]. SAGE tag No. in Spga–Sp cy–Spdt, number of SAGE tags in type A spermatogonia, pachytene spermatocytes, and round spermatids; S, sense; AS, antisense; Mean Δ cycle No. ± SD, means of difference in QPCR cycle number between sense and antisense transcripts and standard deviation of the means with testicular total RNA as template; Approximate S/AS, approximate ratio of expression levels of sense/antisense transcripts calculated from QPCR cycle numbers.
Molecular cloning and characterization of antisense transcripts

*T-complex-associated testis expressed 3*  
To understand further the nature of the antisense transcripts, we cloned and characterized the antisense transcripts of one member each of group 1B, group 2, and group 3. Cloning of the antisense transcripts of *Tcte3* of group 1B yielded 18 distinct clones, which could be assigned to two groups based on the size of the transcript. Fig. 3A shows a cartoon of the *Tcte3* antisense transcripts, *A-Tcte3-a* (GenBank Accession No. DQ072383) and *A-Tcte3-b* (GenBank Accession No. DQ072384). The presence of the longer antisense transcript, *A-Tcte3-b*, was confirmed by amplifying the entire transcript using the most 5′ and 3′ primers in one RT-PCR. The nucleotide sequences of *A-Tcte3-a* and *A-Tcte3-b* are shown in Fig. 3B. *A-Tcte3-a* was identical to the first 532 nucleotides of *A-Tcte3-b* except for an additional 18 A’s. Blasting *A-Tcte3-b* against the mouse genome revealed that the first 704 nucleotides of *A-Tcte3-b* were the reverse complement of exon 4 (164 bp) of *Tcte3* (NT_039641), including 246 and 294 bp, respectively, on the 5′ and 3′ side of this exon (Fig. 3B). However, the 50 nucleotides upstream of the poly(A) tail could not be found in the *Tcte3* gene. These 50

![Fig. 1. Grouping of antisense amplicons based on their structure. For groups 1A, 1B, and 2, antisense amplicons are represented by yellow and brown bars and are shown on the top. Sense gene sequences are represented by red and pink bars and are shown on the bottom. For group 3, pseudogene is represented by a striped bar. The noncoding strand of the sense gene is represented by a white bar. Complementarity is represented by short black lines linking the antisense amplicon with the sense sequence. Thick black lines represent introns.](image-url)
nucleotides contained a polymorphic AAAAC repeat, with seven repeats in some clones and eight repeats in the others interrupted by an AAAC sequence. The SAGE tag sequences identified in the germ cell SAGE libraries were present in the antisense transcripts. Computer-assisted translation of the antisense transcripts revealed open reading frames (ORF) of 117 to 144 bp. The largest ORF encoded a putative polypeptide of 48 amino acids with no recognizable protein motif, indicating that the antisense transcript might serve as a noncoding regulatory RNA [28]. Unlike the antisense amplicon identified by orientation-specific RT-PCR, neither of the Tcte3 antisense transcripts overlapped more than one exon. The cloning approaches selected transcripts with a poly(A) sequence at the 3′ end (3′ rapid amplification of cDNA ends (RACE)) or a cap at the 5′ end (5′RACE). Therefore the results suggested there exist more than two species of antisense transcripts of Tcte3, some of which were not polyadenylated and capped, similar to that observed in a number of noncoding regulatory RNAs [28].

Protamin 1 and 2

Three full-length antisense transcripts (A-Prm[1-2]-a (GenBank Accession No. DQ072380), A-Prm[1-2]-b (GenBank Accession No. DQ072381), and A-Prm[1-2]-c (GenBank Accession No. DQ072383)) of the sole member of group 2, i.e., Prm1, were cloned; upon analysis, these turned out to overlap Prm1 as well as the neighboring Prm2. These transcripts were alternative splice variants containing different numbers of exons. The presence of these transcripts was confirmed by single RT-PCRs using the most 3′ and 5′ primers. The structure of the antisense transcripts is shown in Fig. 4A. The nucleotide sequence of A-Prm[1-2] is shown in Fig. 4B. The shortest antisense transcript was A-Prm[1-2]-a, which had 921 bp and a poly(A) tail. It had four exons, with the first exon overlapping the most 3′ 126 bp of Prm2 (BC049612), and exons 2 and 3 were the reverse complement of Prm1 (NM_013637). A-Prm[1-2]-b was identical to A-Prm[1-2]-a with an additional exon (exon 3) of 116 bp, which overlapped part of the intergenic sequence between Prm1 and Prm2. It had a poly(A) tail. The longest antisense transcript was A-Prm[1-2]-c, which was identical to A-Prm[1-2]-b with an additional exon (exon 2) of 147 bp and a poly(A) tail. The 5′ end of exon 2 overlapped the 5′-most 112 bp of exon 1 of Prm2. Computer-assisted translation of the three antisense transcripts identified a similar ORF of 318 bp encoding a putative polypeptide containing four casein kinase II phosphorylation sites and four myristoylation sites as shown in Fig. 4B. A number of protein-encoding antisense transcripts had been reported previously [20,23,29,30]. Whether the Prm[1-2] antisense transcripts encode any protein awaits validation.

Fig. 2. Nucleotide sequence of group 2 and group 3 antisense amplicons. Primers for amplification are represented by broken line with arrow. Notation of the primers refers to the sense transcript. (A) Prm1-ASF1. The antisense amplicon sequence is on the top. The complementary sense sequence is on the bottom. Uppercase letters represent exon sequence and lowercase letters represent intronic sequence. (B) Calm2-ASF1. The antisense amplicon sequence is on the top. The complementary sense sequence is on the bottom. Uppercase letters represent exon sequence and lowercase letters represent intronic sequence. (C) Ch10-ASF1, antisense amplicon of chaperonin 10. (D) Uba52-ASF1, antisense amplicon of ubiquitin A-52 ribosomal protein fusion product 1. (E) Ubb-ASF1, antisense amplicon of ubiquitin B.
follows rules similar to those of the sense transcripts. It was interesting to note that the Prm\[1-2\] antisense transcripts spanned two neighboring genes. Thus, they might interact with both sense transcripts. Examination of the SAGE tags in the germ cell libraries indicated two antisense tags that overlapped with the second exon of Prm2 [12]. Since SAGE tags are close to the 3′ end of transcripts, this observation suggested the presence of antisense transcripts that overlapped primarily with the Prm2 gene.

Ubiquitin A-52 residue ribosomal protein fusion product 1

The antisense transcripts of Uba52 in group 3 were cloned. 5′ RACE yielded two groups of clones, neither of which were derived from the functional Uba52 gene on chromosome 8 (Fig. 5A). They were derived from two putative pseudogenes, one on the tip of the short arm of chromosome 4 (NT_039258) and the other on the long arm of chromosome 9 (NT_039472). Both pseudogenes were 97–98% homologous to Uba52 mRNA (BC014772) with no introns and a putative ancestral poly(A) stretch with 24 and 17 A’s, respectively. 3′ RACE using primers based on the 5′ RACE products subsequently yielded three groups of antisense transcripts derived from the pseudogene on chromosome 4 (A-Uba52-4a (GenBank Accession No. DQ072385), A-Uba52-4b (GenBank Accession No. DQ072386), and A-Uba52-4c (GenBank Accession No. DQ072387)) and four groups of antisense transcripts derived from the pseudogene on chromosome 9 (A-Uba52-9a (GenBank Accession No. DQ072388), A-Uba52-9b (GenBank Accession No. DQ072389), A-Uba52-9c (GenBank Accession No. DQ072390), and A-Uba52-9d (GenBank Accession No. DQ072391)). The presence of these transcripts was confirmed by successful RT-PCR with the most 5′ and most 3′ primers. Fig. 5A shows the cartoon depicting the intron–exon structure of the Uba52 antisense transcripts. The nucleotide sequences of the antisense transcripts of Uba52 are shown in Fig. 5B. It is interesting to notice that the short amplicon generated by orientation-specific RT-PCR was derived from another pseudogene of Uba52 on chromosome 11, indicating that only a portion of the antisense transcripts of this gene were cloned in this exercise.

All the antisense transcripts derived from the chromosome 4 pseudogene predict an extra exon downstream of the 3′ end of the pseudogene identified by the putative ancestral poly(A) tail (Fig. 5A, b). A-Uba52-4a comprised two exons of 170 and 415
bp, respectively. The first exon was complementary to a putative exon 1830 bp downstream of the pseudogene. This exon was spliced to an exon starting at 84 bp 5′ of the poly(A) stretch of the pseudogene. Splicing of the exons of A-Uba52-4a followed the AG/GT(C) rule [18]. A-Uba52-4b, with 1048 bp, was the largest antisense transcript derived from this pseudogene. It was

Fig. 4. Antisense transcripts of protamine 1 and protamine 2. (A) Cartoon of the Prm[1-2] antisense transcripts. The putative exons of the antisense transcripts are shown. The exons are numbered starting from the 5′ end of the antisense transcript. The rest of the notations are the same as described in the legend to Fig. 3B. The intergenic sequence between the Prm1 gene and the Prm2 gene is ∼4.2 kb. (B) Nucleotide sequence and computer-predicted amino acid sequence of the polypeptide encoded by A-Prm[1-2]-c. A-Prm[1-2]-a does not have exons 2 and 3; A-Prm[1-2]-b does not have exon 2. Exon–intron boundaries are indicated by ←→. The SAGE tag sequences identified in germ cell SAGE libraries are underlined. Amino acids encoded by the ORF are shown underneath the nucleotide sequence as single-letter code. The stop codon is represented by +++. Amino acids constituting the potential casein kinase II phosphorylation sites in the putative encoded polypeptide are underlined. Amino acids constituting potential myristoylation site are marked by asterisks.
similar to A-Uba52-4a except the second exon was larger, with 854 bp. A-Uba52-4c was different from the other two transcripts in that the first exon was 321 bp and was 1799 bp downstream of the pseudogene. This exon was spliced to the second exon at the same position as the other two antisense transcripts. Computer-assisted translation of A-Uba52-4b revealed an ORF of 393 bp with no recognizable protein motif.

The longest antisense transcript derived from the pseudogene on chromosome 9, A-Uba52-9c, had 1920 bp with 298 bp in exon 1, 161 bp in exon 2, and 1461 bp in exon 3, including a poly(A) tail of ~21 A’s. Exon 3 overlapped with the pseudogene of Uba52. Exon 1 was predicted to be 4060 bp and exon 2 to be 1868 bp, downstream of the ancestral poly(A) stretch of the pseudogene (Fig. 5A, c). A-Uba52-9b was identical to A-Uba52-9c except that exon 3 was shorter with only 725 bp and a poly(A) tail of ~14 A’s. A-Uba52-9a did not have exon 2. A-Uba52-9d was identical to A-Uba52-9a except with a LINE 1 repeat of 88 bp inserted between exons 1 and 2. 3’ extension of exon 3 of both A-Uba52-9a and A-Uba52-9d was not successful and we sequenced only 170 bp from the 3’ end and stopped at the poly(T) stretch. Different from A-Uba52-4b, no ORF of appreciable size could be identified in any of the chromosome 9 antisense transcripts.

The study of the antisense transcripts of Uba52 showed that unlike the pseudogenes, the antisense transcripts derived from them could be composed of multiple exons and undergo posttranscription processing. It is likely that these transcripts have biological activities, unlike pseudogenes that are presumed to be nonfunctional. A-Uba52-4b had an appreciable ORF. Even though no known motif could be identified, the potential for it to encode a novel protein could not be precluded. The other transcripts with no significant ORF might have regulatory activities similar to that proposed for noncoding RNAs [28]. The transcription of an antisense RNA from a pseudogene was reported previously for neural nitric oxide synthase [26]. In that case, the pseudogene-derived antisense transcript was shown to regulate the synthesis of the neural nitric oxide synthase protein. Whether the Uba52 antisense transcripts serve as regulatory RNAs is currently under investigation.

Expression studies of antisense transcripts

The relative expression levels in mouse testis of sense and antisense transcripts of the aforementioned 19 genes were examined using QPCR with primer sets described in Supplementary Table S2. The data indicated that the expression level in testis of the sense transcript was always significantly higher (p < 0.001) than that of the antisense transcript. The results are summarized in Table 1. This differential expression of the sense and antisense transcripts in whole testis was similar to that observed in the more differentiated germ cells such as spermatocytes and spermatids as shown by the SAGE analysis, with the exception of Pplic and Sh3-Stam [12] and Table 1. In most cases, the difference in expression levels between the sense and the antisense transcripts as shown by QPCR was much bigger than that revealed by PAGE analysis following RT-PCR (Supplementary Fig. S1). The discrepancy between the QPCR and the PAGE results could be the consequence of differences in the procedure. Since RT-PCR was done using the one-tube procedure while QPCR involved a transfer of RT product from one tube to another prior to performing QPCR, any loss of RT product due to adsorption onto the wall of the RT reaction tube will greatly affect the result of the QPCR but not that of the RT-PCR and PAGE. This effect will be especially significant for low-abundance antisense transcripts. Thus the QPCR results will not accurately reflect the relative expression levels of the sense and antisense transcripts, particularly for the low-abundance transcripts. Another cause of the discrepancy could be the excessive number of PCR cycles (30 cycles) usually performed to reveal the presence of the antisense transcripts.

We also examined the expression of sense and antisense transcripts of nine genes in nontesticular tissues, including whole embryo, ovary, brain, thymus, kidney, spleen, heart, lung, and liver, by semiquantitative orientation-specific RT-PCR. The primers for the RT-PCR were the same as those for studying testicular expression of the sense and the antisense transcripts (Supplementary Table S2). Results of electrophoresis of the RT-PCR products are shown in Fig. 6. The relative expression levels of the sense and the antisense transcripts are summarized in Table 2. Expression of the sense and the antisense transcript of Pdcl2 was testis specific. The expression of the antisense transcript of Tsg1 and Tcte3 was also testis specific. Unlike the sense transcript, the antisense transcript of Pml1 was ubiquitously expressed. Expression of Pm1 and Prm2 in nontesticular tissues was reported previously [33–35] and confirmed by nucleotide sequencing of the TA-cloned sense and antisense amplicons (data not shown). The sense and the antisense transcripts of Pdcl2 and Prm2 were coexpressed in selected tissues while those of Calm2 and Uba52 were coexpressed in all tissues examined. On the other hand, the sense transcripts of Pplic and Pplic were ubiquitously expressed while their antisense transcripts were found only in certain tissues.

Although the sense transcripts were often expressed at higher levels than antisense transcripts, when the two were coexpressed the relative expression pattern of the sense and the antisense transcripts varied for different genes. This was contrary to that observed when mouse embryonic tail genes were studied [36]. Coexpression of the sense and the antisense transcripts in the same cell facilitates the formation of RNA duplexes and may result in modulation of gene expression or mRNA stability and processing [6,37,38]. Tissue-specific expression of the antisense transcripts implies regulation of the appropriate tissue- and cell-type expression of the sense gene similar to that observed for Nphs1 and Dnm3 [37,39,40]. It may also indicate regulation or activity of the antisense transcript independent of the sense transcript as observed in the case of FGF-2 [41].

Differential expression of sense and antisense transcripts in different cells and tissues has been reported for a number of imprinted genes [6,21,42]. Antisense transcripts have been suggested to play important roles in the regulation of
monoallelic expression in X-chromosome inactivation and genomic imprinting [43–45]. A number of the genes in the present study also showed differential expression of the sense–antisense transcripts. The sense transcript of these genes, including Prm1, Prm2, Tcet3, and Tsg1, coincidentally were preferentially expressed in meiotic and postmeiotic cells (Tables 1 and 2). It is unknown whether these antisense transcripts have a particular function in transcription regulation at these stages.
The present study demonstrates the occurrence of antisense transcription in germ cells. It is tempting to speculate that this vigorous control of gene expression is necessary to ensure the accuracy or to facilitate the functioning of the biological processes occurring during spermatogenesis such as genome-wide methylation–demethylation, genomic imprinting, monoallelic gene expression, etc. It suggests that mouse spermatogenesis may be a good model for studying the regulation and biological activities of antisense transcripts.

It has been argued that antisense transcripts arise from leaky transcription of the noncoding strand. Even though this possibility cannot be completely ruled out, the fact that most of the antisense transcripts examined are polyadenylated or processed in a manner similar to the sense transcripts suggests that they are intentional transcripts. This is supported by the results of a recent study comparing the genomic organization of genes with or without antisense transcripts between human, mouse, and pufferfish [46]. Previous studies on antisense transcription either used a computational approach to look for the global presence of antisense transcripts or focused on a single gene [1–4,19–26,29,30,35,37]. Few reports document the mechanisms by which the antisense transcripts are generated. In the present study, similar to a recent report [47], we used a SAGE database for the identification of antisense transcripts. We showed that a significant percentage of differentially expressed genes in spermatogenic germ cells are associated with antisense transcripts. These transcripts come from a wide spectrum of sources: processed sense mRNA, the sense gene locus, pseudogene, two neighboring genes, and intergenic sequence. Characterization of the cloned antisense transcripts also showed that antisense transcripts can be processed, alternatively spliced, capped, and polyadenylated. There are also antisense transcripts that do not have a 5′ cap and/or 3′ poly(A) tail. Thus the origin and structure of antisense transcripts are very complex. Antisense transcripts have been proposed to have functions in transcriptional and posttranscriptional regulation [6,14,28,43,44,48,49]. The complex nature of the antisense transcripts suggests that they are well suited to provide an additional layer of vigorous regulation of gene expression.

Materials and methods

Orientation-specific RT-PCR and QPCR

Presence of antisense transcript was confirmed by orientation-specific RT-PCR. Total RNA from mouse tissues was purchased from Ambion (Austin, TX, USA). RNA integrity and concentration were checked with the Bioanalyzer 2100 (Agilent Technologies, Germantown, MD, USA). Orientation-specific RT-PCR was performed as described previously by Shendure and Church [14]. The One Step RT-PCR kit (Qiagen, Valencia, CA, USA) was used according to the manufacturer’s protocol with a total volume of 25 μl. The primers were designed based on the published mRNA sequence such that they would prime the amplification of a 120- to 230-bp sequence covering the most 3′ sense–antisense pair. Primers were designed using Primer Express version 2.0 (Applied Biosystems, Branchburg, NJ, USA) and specificity was confirmed by a BLASTN search against the nonredundant and EST mouse sets from NCBI. The sequence and location of the primers are described in Supplemental Table S2. One microgram of total RNA was used as template in all RT reactions. In general, 30 PCR cycles were performed for the orientation-specific RT-PCR with the exception of cases in which the expression level of the transcript was very high or very low. All orientation-specific RT-PCR experiments were performed with controls in which no strand-specific primer was added to the RT reaction [14]. Five microliters of the 25 μl product of orientation-specific RT-PCR was analyzed by PAGE. Quantitative real-time RT-PCR analyses were carried out as described previously [50]. Sense-specific first-strand cDNA was synthesized as described earlier using the Qiagen One Step RT-PCR kit. Products of the RT reaction were aliquoted for QPCR. Primers for QPCR were the same as those used for orientation-specific RT-PCR as described in Supplemental Table S2. QPCR was performed in triplicate in the 7900 HTS Sequence Detection System (Applied Biosystems, Foster City, CA, USA). Gene expression data were reported after normalizing to 18S rRNA content. The mean cycle number and the standard deviations were reported. Significance of expression level difference was analyzed using Student’s t test.

5′ and 3′ RACE and nucleotide sequencing

To clone the antisense transcripts, anchored PCR was performed using the GeneRacer Kit (Invitrogen, Gaithersburg, MD, USA) according to the manufacturer’s recommendations. 5′ RACE using the GeneRacer Kit specifically targets only 5′-capped mRNA. Reverse transcription of the ligated mRNA to create RACE-ready first-strand cDNA was effected using Superscript III reverse transcriptase (Invitrogen) and orientation-specific primer. The RACE product was amplified using nested primers and gene-specific primers before TA cloned into the pCR 4-TOPO vector (Invitrogen) according to the manufacturer’s protocol. The sequences of specific primers used for 5′ and 3′ RACE are indicated in Figs. 3B, 4B, and 5B. The sequence of the cloned cDNA was determined by cycle sequencing using BigDye Primer Cycle Sequencing Kits (Applied Biosystems). DNA sequences were analyzed using DNASIS software version 2.5 (MiraBio, Alameda, CA, USA). The presence of antisense transcripts was confirmed by repeating the RT-PCR using the most 5′ and the most 3′ primers that were located in different exons in the antisense transcript. If both primers were contained in the sense transcript, orientation-specific RT-PCR was performed.

Open reading frames in the nucleotide sequence were predicted using the NCBI ORF finder program (http://www.ncbi.nlm.nih.gov/gorf/gorf.html) and ORF Finder (http://www.bioinformatics.org/sms/orf_find.html). Translated amino acid sequences in the potential ORFs in FASTA format were input into the SMART 4.0 database [51]. The output was ranked by e values. Domains with scores that were less significant than the required threshold or overlapped with some other source of annotation were rejected.
Fig. 6. Polyacrylamide gel electrophoresis of the product of orientation-specific RT-PCR of nine genes in different tissues. M, size marker; S, orientation-specific RT-PCR using primer specific for the sense transcript; As, orientation-specific RT-PCR using primer specific for the antisense strand. Control with no strand-specific primer in the RT reaction was done in all experiments. The representative results of triplicate experiments are shown.
Table 2

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Calm2</td>
<td>++++/+++</td>
<td>+++/+</td>
<td>++++/++</td>
<td>++++/+++</td>
<td>+++/+</td>
<td>++++/+++</td>
<td>+++/+</td>
<td>++++/+++</td>
<td>+++/+</td>
<td>+++/+</td>
</tr>
<tr>
<td>Dct12</td>
<td>–/-</td>
<td>++/+</td>
<td>–/-</td>
<td>–/-</td>
<td>–/-</td>
<td>–/-</td>
<td>–/-</td>
<td>–/-</td>
<td>–/-</td>
<td>–/-</td>
</tr>
<tr>
<td>Ppdc1</td>
<td>++++/++</td>
<td>–/-</td>
<td>+++/++</td>
<td>–/-</td>
<td>–/-</td>
<td>–/-</td>
<td>–/-</td>
<td>–/-</td>
<td>–/-</td>
<td>–/-</td>
</tr>
<tr>
<td>Ppdc2</td>
<td>++/++</td>
<td>++/+</td>
<td>++/+</td>
<td>++/+</td>
<td>++/+</td>
<td>++/+</td>
<td>++/+</td>
<td>++/+</td>
<td>++/+</td>
<td>++/+</td>
</tr>
<tr>
<td>Prm1</td>
<td>–/+</td>
<td>+++/+</td>
<td>–/+</td>
<td>–/+</td>
<td>–/+</td>
<td>–/+</td>
<td>–/+</td>
<td>–/+</td>
<td>–/+</td>
<td>–/+</td>
</tr>
<tr>
<td>Prm2</td>
<td>–/-</td>
<td>+++/+</td>
<td>–/-</td>
<td>–/-</td>
<td>–/+</td>
<td>–/+</td>
<td>–/+</td>
<td>–/+</td>
<td>–/+</td>
<td>–/+</td>
</tr>
<tr>
<td>Tcte3</td>
<td>–/-</td>
<td>+++/+</td>
<td>–/-</td>
<td>–/-</td>
<td>–/-</td>
<td>–/-</td>
<td>–/-</td>
<td>–/-</td>
<td>–/-</td>
<td>–/-</td>
</tr>
<tr>
<td>Tsg1</td>
<td>+/+</td>
<td>–/-</td>
<td>–/-</td>
<td>–/-</td>
<td>–/-</td>
<td>–/-</td>
<td>–/-</td>
<td>–/-</td>
<td>–/-</td>
<td>–/-</td>
</tr>
<tr>
<td>Uba52</td>
<td>+++/+</td>
<td>++++/++++++</td>
<td>++++/+++</td>
<td>++++/+++</td>
<td>++++/+++</td>
<td>++++/+++</td>
<td>++++/+++</td>
<td>++++/+++</td>
<td>++++/+++</td>
<td>++++/+++</td>
</tr>
</tbody>
</table>

S, sense transcript expression; AS, antisense transcript expression; + indicates the presence of a band and the number of +’s represents the visual intensity of the band; – indicates the absence of a band.

Acknowledgments

This research was supported by the Intramural Research Program of the NIH, National Institute of Child Health and Human Development.

Appendix A. Supplementary data

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.ygeno.2005.12.006.

References

[34] Stanford Microarray Database [http://www.genome-www5.stanford.edu/cgi-bin/source].