Available online at www.sciencedirect.com

SCIENCE DIRECT

ELSEVIER

Genomics 88 (2006) 127–131

GENOMICS

www.elsevier.com/locate/ygeno

Short Communication

Reverse transcriptase template switching and false alternative transcripts

Julie Cocquet [a,b], Allen Chong [c], Guanglan Zhang [c], Reiner A. Veitia [a,b,*]

[a] *INSERM U709, Hôpital Cochin, Pavillon Baudelocque, 123 Bd de Port Royal, 75014 Paris, France*
[b] *Université Paris 7 Denis Diderot, France*
[c] *Institute for Infocomm Research, 21 Heng Mui Keng Terrace, Singapore 119613, Singapore*

Received 9 November 2005; accepted 20 December 2005
Available online 2 February 2006

Abstract

Reverse transcriptase (RT) can switch from one template to another in a homology-dependent manner. In the study of eukaryotic transcripts, this propensity of RT can produce an artificially deleted cDNA, which can be wrongly interpreted as an alternative transcript. Here, we have investigated the presence of such template-switching artifacts in cDNA databases, by scanning a collection of human splice sites (Information for the Coordinates of Exons, ICE database). We have confirmed several cases at the experimental level. Artifacts represent a significant portion of apparently spliced sequences using noncanonical splice signals but are rare in the context of the whole database. However, care should be taken in the annotation of alternative transcripts, especially when the RT used is poorly thermostable and when the putative intron is flanked by direct repeats, which are the substrate for template switching.

Keywords: Reverse transcriptase; Artifact; Template switching; Splice isoforms; RT-PCR; EST database; cDNA library; FOXL2

During reverse transcription of a retroviral genome, reverse transcriptase (RT) enzyme must undergo two template-switching events, which occur at the ends of the template, to complete DNA synthesis [1]. RT undergoes frequent additional switches of templates, leading to a great variability in retroviral populations [2]. These template-switching (TS) events are generated in a homology-dependent manner. Specifically, the growing DNA strand synthesized by the RT dissociates from its original template and reassociates with another template, presenting a homologous region. When it involves two distinct templates, this process is called intermolecular TS. When the RNA template contains direct repeats, the switch can occur from one portion to another of the same RNA molecule and therefore lead to a deletion of the intervening region. This process is called intramolecular TS.

Intermolecular template switching of the RT can produce chimeric cDNAs [3], whereas intramolecular TS can lead to deleted cDNA forms, which can be misinterpreted as splice isoforms (cf. Fig. 1A). For instance, the *Nox1* gene short transcript (NOH-1S) results from an intramolecular template-switching artifact caused by the presence of two direct repeats surrounding the deleted region [4]. A deletion within the ribonucleotide reductase transcript has also been shown to be an artifact induced by the formation a stem loop structure, potentially bringing into vicinity the two repeats flanking the deleted region [5]. Mader et al. observed that large distances between direct repeats increase the frequency of template switching [5], as previously demonstrated by Delviks and Pathak [6]. Furthermore, cDNA products are mainly studied by polymerase-chain-reaction (PCR) experiments or cDNA library screening. Both techniques bias the representation in favor of shorter transcript forms (arising or not from TS).

In our study of the *FOXL2* gene, considered so far as intronless [7], we have observed an unexpected shorter cDNA form, potentially generated by a similar mechanism. Indeed, reverse transcription followed by PCR (RT-PCR) on human ovary RNA expressing *FOXL2* produced two *FOXL2* cDNA forms: the longest corresponding to the known transcript, whereas the other possessed a small (77 nt) internal deletion (Fig. 1B). This putative cDNA produced a signal as intense as the one for a full-length

* Corresponding author. INSERM U709, Hôpital Cochin, Pavillon Baudelocque, 123 Bd de Port Royal, 75014 Paris, France. Fax: +33 1 43 26 44 08.
E-mail address: veitia@cochin.inserm.fr (R.A. Veitia).

0888-7543/$ - see front matter © 2005 Elsevier Inc. All rights reserved.
doi:10.1016/j.ygeno.2005.12.013

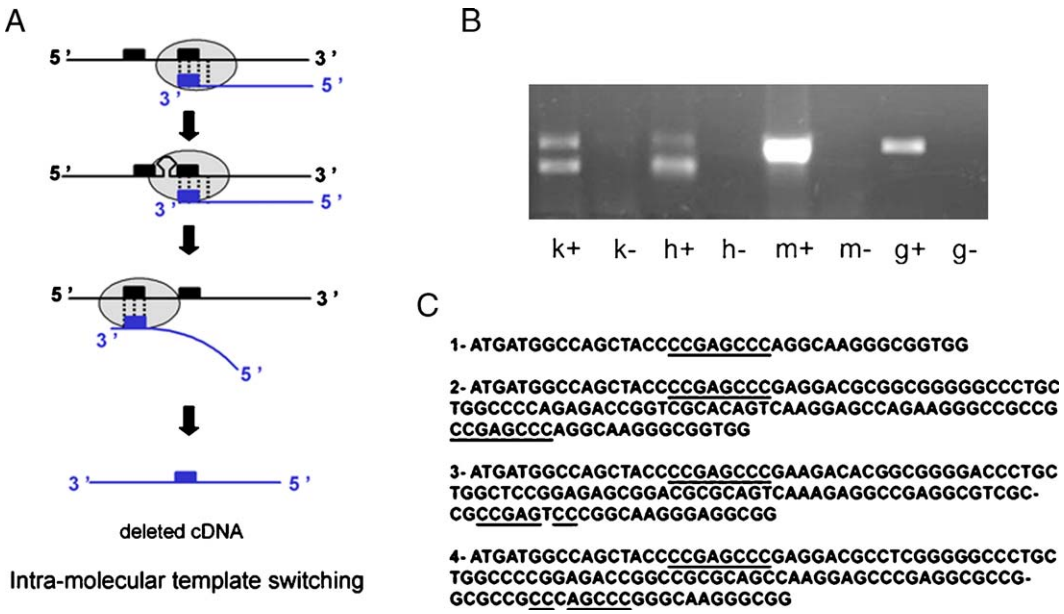

Fig. 1. (A) Schematic representation of intramolecular template switching of the reverse transcriptase. The black line represents RNA molecules whereas the blue line represents a growing cDNA molecule. Boxes represent direct repeats on which the RT enzyme (light gray sphere) may switch template. An RNA with two direct repeats is a potential template for a switching of the RT from one repeat to the other. This may be enhanced by a secondary structure bringing into vicinity the repeats. The resulting cDNA molecule possesses only one copy of the direct repeats. (B) RT-PCR to detect *FOXL2*. k, RNA from KGN cell line (human ovarian granulosa-like tumor cell [12]); h, human ovary total RNA (Ambion); m, mouse ovarian RNA; g, goat ovarian RNA. **+/–**: presence/absence of reverse transcriptase. RT-PCR on human RNA (ovary or KGN cell line) produces two bands, whereas only the longest band is amplified on mouse or goat RNA. (C) 1, Partial sequence of the "deleted" human *FOXL2* cDNA. The 8-nt region (underlined) is present only once. 2, Sequence of the corresponding genomic region of human *FOXL2*. The direct repeats flanking the deleted region are underlined. Note that there are no canonical splice sites at the boundaries of the deletion. 3 and 4, In mouse and goat (respectively) the first 8-nt region (underlined) is conserved but not the second one (broken line).

transcript. The short "transcript" could either result from an uncharacterized splice event or be an artifact of the RT reaction.

In the vast majority of cases, introns start with a GT at the donor splice site and end with an AG acceptor site [8]. However, more rarely, RNA splicing occurs at GC–AG, AT–AC, or GT–GG donor–acceptor pairs of sites. Altogether, these four pairs account for 99.16% of all cases [9]. The flanking regions of the deletion in the shorter *FOXL2* cDNA described here did not encompass any of the known splice sites but involved 8-nt-long direct repeats. This 8-nt sequence was found

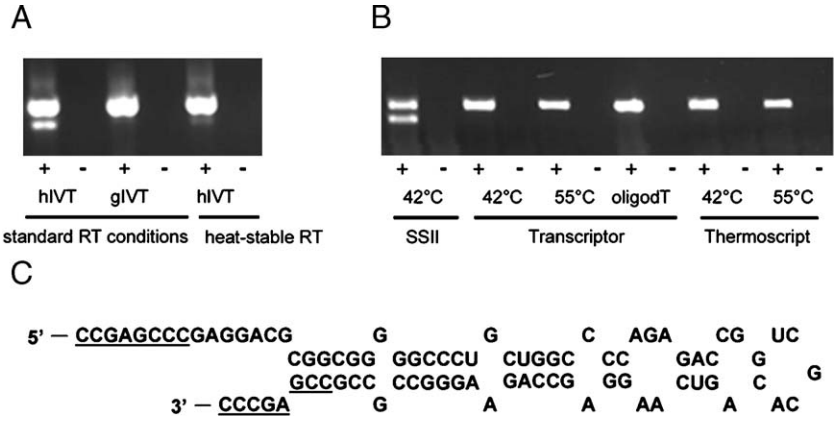

Fig. 2. (A) RT-PCR on human or goat in vitro transcribed *FOXL2* RNA (hIVT and gIVT, respectively). Standard conditions of RT correspond to a primer-specific reverse transcription using Superscript II enzyme at 42°C. The heat-stable Transcriptor enzyme (Roche) was alternatively used under the same conditions but at 55°C. All PCR were performed using FOXL2-A and GFPR-2 primers (online Table 2). RT-PCR with Superscript II RT on hIVT produced two bands, whereas only the longest band was amplified using a heat-stable RT. Similar experiments using goat IVT RNA only produced the longest band. These results indicate that the small band is an artifact of the Superscript II RT. (B) *FOXL2* RT-PCR on KGN total RNA using different reverse-transcription conditions. As RT enzyme, we used either a standard RT (Superscript II, SSII) or a thermostable RT (Transcriptor or Thermoscript). Reverse transcriptions were performed at 42 or 55°C. A primer-specific RT was performed in all cases except one, where an oligo(dT) RT was carried out at 55°C using Transcriptor RT (lane "oligodT"). **+/–**: presence/absence of transcriptase. Primers: FOXL2-A and 395R. The small band corresponds to an artifact obtained when using a poorly heat-stable RT. (C) Potential RNA secondary structure of the region deleted in *FOXL2* artifactual cDNA. The 8-nt direct repeats, which flank the deleted region, are underlined (MFOLD at http://bioweb.pasteur.fr/seqanal/interfaces/mfold-simple.html, default parameters for RNA).

only once on the cDNA sequence (Fig. 1C). To determine whether this deletion was a true splicing event or an artifact, we cloned a nonspliced fragment of the corresponding human region (NM023067: 238–516) to perform in vitro transcription. In vitro transcribed (IVT) RNA molecules were used as template for RT-PCR. As we observed the additional smaller band after RT-PCR performed under the same conditions, we concluded that this cDNA was an experimental artifact (Fig. 2A, standard RT conditions). Interestingly, in other species, as in mouse (*Mus musculus*) or goat (*Capra Hircus*), this artifact was not observed (Fig. 1B). Indeed, the corresponding genomic sequence in mouse and goat did possess the first 8-nt motif as in human; however, the second differed by one nucleotide change (Fig. 1C and Fig. 2A, standard RT conditions).

Our previous experiments had been performed with the Superscript II RT, an engineered version of the Moloney murine leukemia virus RT (M-MLV) (Superscript II, Invitrogen Life Technologies). This enzyme possesses a reduced ribonuclease H (RNaseH) activity and increased thermal stability (optimal temperature 42°C). However, when using more heat-stable RTs (Thermoscript, Invitrogen Life Technologies or Transcriptor, Roche diagnostics) on in vitro transcribed (IVT) *FOXL2* RNA as well as on human ovary total RNA, the artifactual band was not detected anymore (Fig. 2A, heat-stable RT, and Fig. 2B). This was observed at both 55°C, the working temperature of these enzymes, and at 42°C. Similar results were obtained with a gene-specific or an oligodeoxythymidine (oligo(dT)) reverse transcription (Fig. 2B). This confirms the artifactual nature of the above-described *FOXL2* cDNA form. A secondary structure predicted to form between the two direct repeats of *FOXL2* RNA molecule might explain the high frequency of this template switching (Fig. 2C).

The capacity of M-MLV to switch from one template to another (i.e., intermolecular TS) has been shown to increase at a higher incubation temperature. The presence of RNase H activity associated with the RT, as well as the length of the sequence identity between the two templates, has also been shown to modulate the rate of TS [10]. Unlike the intermolecular phenomenon, the intramolecular TS observed here does not seem to be influenced by the presence of RNase H activity or by a high incubation temperature, when using heat-stable RTs. Indeed, both heat-stable RTs used did not display any detectable template switching, whatever the incubation temperature, even if Transcriptor RT did possess RNase H activity (Fig. 2B).

The use of standard RTs (or RTs with increased thermostability, such as Superscript II) in RT-PCR experiments may lead to the production of false alternative transcripts. Moreover, these RTs are frequently used to produce cDNA libraries [11]. The presence of such intramolecular template-switching artifacts in cDNA databases may be misinterpreted as splice events involving noncanonical or canonical splice sites. In order to test this possibility, we first concentrated on noncanonical splice events (i.e., involving donor–acceptor pairs other than GT–AG) and analyzed the

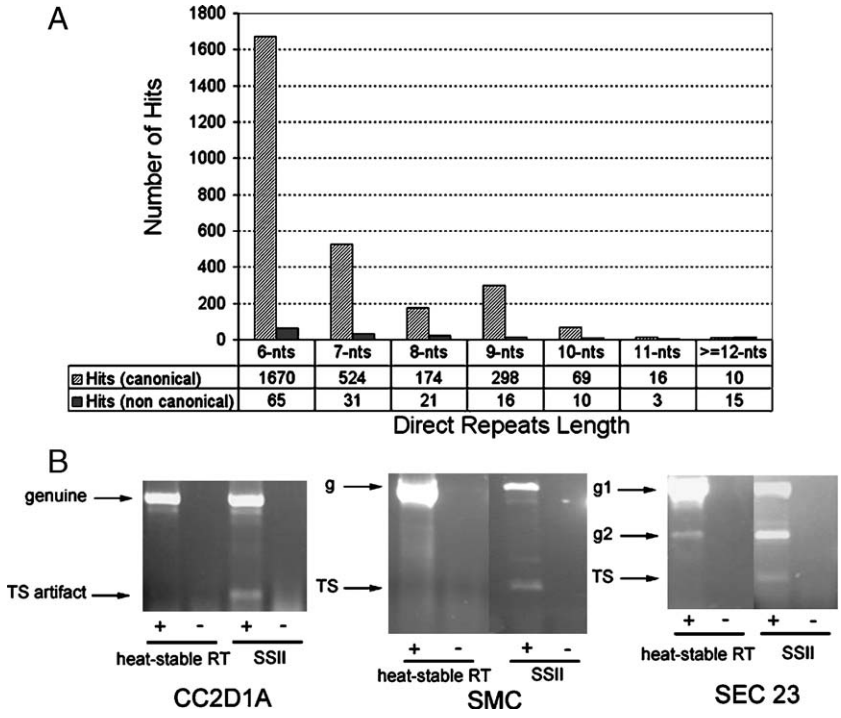

Fig. 3. (A) Number of transcripts, with canonical or noncanonical donor–acceptor sites, presenting a repeated sequence at the boundaries of their intron (from the human splice sites database [9]). They are classified according to the length of the repeated sequence (i.e., 6, 7, 8, 9, 10, 11, or ≥12 nt). (B) RT-PCR on template-switching examples found in the noncanonical splice sites database (additional information in online Table 1). Primer-specific reverse transcriptions were performed on RNA extracted from HeLa or KGN cell lines using various transcriptases (heat-stable RT, Transcriptor at 55°C; SSII, Superscript II at 42°C). The resulting cDNA were PCR-amplified for 34 cycles (primers in online Table 2). +/−: presence/absence of reverse transcriptase. On the left, the 832-bp band corresponding to the CC2D1A genuine transcript is amplified. A smaller (artifactual) band is amplified when using Superscript RT but not when using the Transcriptor RT. Similar results were obtained when performing various RT-PCR experiments on SEC23IP and SMC2 transcripts. g1, g2: genuine transcripts.

boundaries of these noncanonically spliced transcripts present in the largest human splice sites database (Information for the Coordinates of Exons, ICE) [9]. We searched for the presence of direct repeats of 6 to 12 nt in the regions immediately upstream and downstream of the potential splice sites. For 9- to 12-nt repeated sequences, one mismatch was allowed within the sequence. Out of the 1617 noncanonical donor–acceptor pairs present in the ICE database [9], 161 pairs were found to fulfill these criteria, 65 of which displayed direct repeats of 8 nt or more (Fig. 3A). Next, we focused on the examples involving 8-nt or more in direct repeats. In most of these 65 cases, the so-called noncanonically spliced sequence corresponded to one or several EST, whereas the corresponding bona fide transcripts (RefSeq RNAs, i.e., accession NM_000000 or XM_000000) in GenBank were not spliced in the relevant region (Supplemental information). Consequently, these "bona fide" transcripts possessed the two direct repeats and could serve as template for the switching. Among the 65 potential examples of template switching with direct repeats of a minimum 8-nt, 9 were studied by RT-PCR to determine whether they were true splice isoforms or RT artifacts (see online Table 1). We performed primer-specific or oligo(dT) reverse transcription on total RNA extracted from two cell lines (HeLa and KGN), using standard (M-MLV, Superscript II) and heat-stable (Thermoscript, Transcriptor) RTs. PCR was then performed on the resulting cDNA using primers designed to amplify both the noncanonically spliced cDNAs and the longer bona fide corresponding cDNAs. In 3 cases, the noncanonically "spliced" isoform was visible after RT-PCR with standard RTs but not with heat-stable RTs, suggesting that they were RT artifacts (Fig. 3B and online Table 1). For the 6 other cases, only the amplification corresponding to the genuine (longer) transcript was observed, suggesting either that the spliced isoform was not expressed in the cell lines or less likely that template switching was undetectable by PCR (yet it was cloned in the cDNA library). Consequently, these results suggest the presence of template-switching artifacts in cDNA databases. There were no differences in the length of the words between the 3 demonstrated RT artifacts and the 6 others. However, the distance between the two direct repeats (also called spacer) in the genuine sequence was longer in the 3 TS examples (spacer > 640 nt) than in the 6 others (range: 150 nt < spacer < 550 nt, $P = 0.019$). This observation suggests that longer intervening sequences increase the probability to observe a template-switching artifact. However, a template switching can be also detectable when the spacer is short, particularly if it forms a strong secondary structure, as observed for *FOXL2*.

We then searched for a similar phenomenon in transcripts involving canonical splice sites. Out of 90,000 donor–acceptor pairs, 2761 cases contained direct repeats of 6-nt or more at the boundaries of the putative intron and 393 presented repeats of ≥8 nt (Fig. 3A). However, most of the canonical examples (with ≥8-nt direct repeats) that we found matched the majority (if not all) of the corresponding ESTs, indicating that they were genuine spliced transcripts. Besides, we could not identify in silico longer alternative transcripts carrying two direct repeats which could have served as template for switching. Accordingly, when performing the comparative reverse-transcription experiments for 3 of them, the potential TS was amplified whatever the RT used, confirming that they were genuine spliced transcripts. These observations suggest that template-switching artifacts are rare among transcripts with canonical splice sites.

In summary, we suggest that when transcripts carry direct repeats as short as 8 nucleotides, the homology-dependent template-switching property of RTs, especially of poorly thermostable enzymes, is likely to generate products which can be misinterpreted as alternative splice isoforms. Depending on the secondary structure of the transcript, intramolecular template switching can occur at high efficiency. It does not seem to be affected by the incubation temperature or the presence/absence of an associated RNase H activity but might be more efficient when the distance between the direct repeats increases. We suggest that such RT artifacts are present in cDNA databases and that they may represent a relevant portion of the noncanonically spliced sequences but we cannot exclude that some template-switching artifacts are misinterpreted as canonically spliced transcripts. To circumvent these artifacts, care should be taken when performing RT-PCR and when generating cDNA libraries. We recommend the use of thermostable RTs and to be cautious when identifying a splice variant that displays direct repeats around the boundaries of the putative intron, especially when involving noncanonical splice sites.

Acknowledgments

We thank Lara Moumné and Sandrine Caburet for reading the manuscript. J.C. is funded by Fondation Recherche Médicale (FRM). R.A.V. is funded by the Université Paris 7, the INSERM, and a GIS grant.

Appendix A. Supplementary data

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.ygeno.2005.12.013.

References

[1] E. Gilboa, S.W. Mitra, S. Goff, D. Baltimore, A detailed model of reverse transcription and tests of crucial aspects, Cell (1979) 93–100.
[2] H.M. Temin, Retrovirus variation and reverse transcription: abnormal strand transfers result in retrovirus genetic variation, Proc. Natl. Acad. Sci. USA (1993) 6900–6903.
[3] X.C. Zeng, S.X. Wang, Evidence that BmTXK beta-BmKCT cDNA from Chinese scorpion Buthus martensii Karsch is an artifact generated in the reverse transcription process, FEBS Lett. (2002) 183–184.
[4] M. Geiszt, K. Lekstrom, T.L. Leto, Analysis of mRNA transcripts from the NAD(P)H oxidase 1 (Nox1) gene. Evidence against production of the NADPH oxidase homolog-1 short (NOH-1S) transcript variant, J. Biol. Chem. (2004) 51661–51668.
[5] R.M. Mader, et al., Reverse transcriptase template switching during reverse transcriptase-polymerase chain reaction: artificial generation of deletions in ribonucleotide reductase mRNA, J. Lab. Clin. Med. (2001) 422–428.

[6] K.A. Delviks, V.K. Pathak, Effect of distance between homologous sequences and 3′ homology on the frequency of retroviral reverse transcriptase template switching, J. Virol. (1999) 7923–7932.
[7] J. Cocquet, et al., Structure, evolution and expression of the FOXL2 transcription unit, Cytogenet Genome Res. (2003) 206–211.
[8] T. Strachan, A.P. Read, Hum. Mol. Genet. 3 (2004) 19–21.
[9] A. Chong, G. Zhang, V.B. Bajic, Information for the Coordinates of Exons (ICE): a human splice sites database, Genomics (2004) 762–766.
[10] G.X. Luo, J. Taylor, Template switching by reverse transcriptase during DNA synthesis, J. Virol. (1990) 4321–4328.
[11] D.S. Gerhard, et al., The status, quality, and expansion of the NIH full-length cDNA project: the Mammalian Gene Collection (MGC), Genome Res. (2004) 2121–2127.
[12] Y. Nishi, et al., Establishment and characterization of a steroidogenic human granulosa-like tumor cell line, KGN, that expresses functional follicle-stimulating hormone receptor, Endocrinology (2001) 437–445.