Genomics 96 (2010) 333–341

Contents lists available at ScienceDirect

Genomics

journal homepage: www.elsevier.com/locate/ygeno

Genome-wide analysis of expression modes and DNA methylation status at sense–antisense transcript loci in mouse

Yutaka Watanabe [a,1,2], Koji Numata [a,1], Shinya Murata [c,d], Yuko Osada [c,d], Rintaro Saito [c,d], Hajime Nakaoka [e], Naoyuki Yamamoto [e], Kazufumi Watanabe [f], Hidemasa Kato [g], Kuniya Abe [a,b], Hidenori Kiyosawa [a,b,*]

[a] Technology and Development Team for Mammalian Cellular Dynamics, BioResource Center (BRC), RIKEN Tsukuba Institute, Ibaraki 305-0074, Japan
[b] Graduate School of Life and Environmental Sciences, University of Tsukuba, Ibaraki 305-0006, Japan
[c] Institute for Advanced Biosciences, Keio University, Tsuruoka 997-0017, Japan
[d] Faculty of Environment and Information Studies, Keio University, Fujisawa, 252-8520, Japan
[e] C's Lab Co. Ltd., Maruito Sapporo Bldg., 7F, Kita 2 Nishi 1, Kita-ku, Sapporo, 060-0002, Japan
[f] Custom Biotechnology Service Group, Hokkaido System Science Co. Ltd. 2-1 Shinkawa Nishi 2-1, Kita-ku, Sapporo, 001-0932, Japan
[g] Research Center for Genomic Medicine, Saitama Medical University, Saitama 350-1241, Japan

ARTICLE INFO

Article history:
Received 7 December 2009
Accepted 18 August 2010
Available online 22 August 2010

Keywords:
Sense–antisense transcripts
Gene expression
DNA methylation
MeDIP chip
Mouse
Microarray

ABSTRACT

The functionality of sense–antisense transcripts (SATs), although widespread throughout the mammalian genome, is largely unknown. Here, we analyzed the SATs expression and its associated promoter DNA methylation status by surveying 12 tissues of mice to gain insights into the relationship between expression and DNA methylation of SATs. We have found that sense and antisense expression positively correlate in most tissues. However, in some SATs with tissue-specific expression, the expression level of a transcript from a CpG island-bearing promoter is low when the promoter DNA methylation is present. In these circumstances, the expression level of its opposite-strand transcript, especially when it is poly(A)-negative was coincidentally higher. These observations suggest that, albeit the general tendency of sense–antisense simultaneous expression, some antisense transcripts have coordinated expression with its counterpart sense gene promoter methylation. This cross-strand relationship is not a privilege of imprinted genes but seems to occur widely in SATs.

1. Introduction

Sense–antisense transcripts (SATs) are pairs of transcriptional units that fully or partially overlap and are transcribed in opposite directions (for a review, see [1]). To analyze transcripts as SAT units would provide a deep understanding of gene regulatory network because some functional relationships between sense and antisense transcripts were inferred in some SATs [2–5]. Our previous studies demonstrated that 1) a substantial number of SATs [291 of 1486 (~20%) SATs in humans and 1948 (~15%) of those in mice] are conserved evolutionarily, 2) about 33% of the conserved pairs showed similar expression patterns between human and mouse [6], 3) SATs are found less frequently on the X chromosome in human and mouse [7], and 4) for approximately 60% of SATs in mice, the sense:antisense expression ratio fluctuated among tissues and cell types [6,8]. These observations imply that SATs play a role in the maintenance of tissue-specific gene expression programs and/or possibly in cell differentiation.

We also previously reported that SATs tend to lack 3′ polyadenylation (poly(A)-negative RNA) and are localized to the nucleus [8]. Although poly (A)-negative transcripts are known to be abundantly expressed in mammalian cells [9], transcriptome analyses have focused mainly on poly(A)-positive transcripts. This approach may need to be revised in view of a recent report showing that poly(A)-negative transcripts potentially comprise almost half of the human transcriptome [10]. It has been assumed that many non-protein-coding (npc) transcripts are not polyadenylated and are localized to the nucleus [11]. We, as well as others, have shown that many npc transcripts can be detected originating from SAT loci [6,12]. Although the number of tissues analyzed was limited we also observed tissue-specific expression of potential poly(A)-negative transcripts from SAT loci.

The expression of some tissue-specific genes correlates with DNA methylation status, which alters during development. DNA methylation is generally considered as a repressive epigenetic modification and is critical for the control of gene expression in mammals. Mice lacking the DNA methyltransferase that catalyses DNA methylation die early during development [13,14]. The status of DNA methylation varies among cell types and developmental time points [15]. DNA methylation is

* Corresponding author. Present address: Transdisciplinary Research Integration Center, National Institute of Genetics, Mishima, Shizuoka, 411–8540, Japan.
E-mail address: hkiyosaw@lab.nig.ac.jp (H. Kiyosawa).
[1] These authors contributed equally to this work.
[2] Present address: Ecology and Evolution Unit, Okinawa Institute of Science and Technology, 1919-1 Tancha, Onna-son, Kunigami, Okinawa 904-0412, Japan.

0888-7543/$ – see front matter © 2010 Elsevier Inc. All rights reserved.
doi:10.1016/j.ygeno.2010.08.007

particularly involved in a number of processes including cell fate determination at various developmental stages [16], monoallelic expression of imprinted genes [17], and transposon silencing [18]. The tissue-specific differentially methylated regions (tDMRs), which may confer tissue specificity to nearby genes, have been identified by genome-wide methylation analyses [19–22]. In fact, tDMRs are observed in sequences up to 2 kb from CpG islands (CGI), regions that are termed "CGI shores." The tDMR methylation status of CGI shores correlates strongly with gene expression [23]. Several studies have proposed that natural antisense transcripts play a role in epigenetic modifications including DNA methylation of the sense transcripts [3,24]. For example, antisense transcription at the α2-globin gene (HBA2) in humans has been proposed to result in hypermethylation of a CGI within the promoter of the α2-globin gene that causes α-thalassemia [24].

Therefore, we analyzed the correlation between the expression pattern of both poly(A)-positive and poly(A)-negative transcripts of SATs as well as the DNA methylation status of SAT loci on a genome-wide scale. Since the DNA methylation status of most SAT loci has not yet been analyzed, we were unable to directly compare SAT expression and DNA methylation profiles by using published data alone. To resolve this problem we used "MeDIP chip" technology, whereby immunoprecipitated methylated DNA fragments can be interrogated directly against microarray platforms [19]. This analysis implies an important role of a subset of npc transcripts in stabilizing a transcriptional silencing on its opposite strand in concert with DNA methylation. Our genome-wide analyses revealed that bidirectional SAT signals are simultaneously detected in most adult tissues. However, a subset of npc transcripts which is expressed in a tissue-specific manner, was not accompanied by its SAT counterpart, which in turn, was likely to be reciprocally silenced by promoter DNA methylation.

2. Results

2.1. Many SATs are simultaneously expressed in the same tissue

The definition of "sense" and "antisense" has been largely arbitrary. For example, scientists of the genomic imprinting field commonly use these terminologies to discriminate "protein-coding (pc)" transcripts (=sense) versus their opposite-strand transcripts, which are vastly "non-protein-coding (npc)." However, in the current study, we have not "fixed" the usage of these terms. As a matter of fact, a large number of SAT pairs consist of two npc transcripts. Therefore, in this current study, we will independently assess their protein-coding capabilities. Hence, whenever these words are originally used in this text, we will simply imply to their "relativeness", which is that they are to be mutually transcribed in the opposite genomic directions with certain degree of overlap.

We have reported previously that two types of pairing arrangement are observed for SATs [25]. In one type, the exons of each transcript fully or partially overlap (Fig. S1A–C) and in the other type, the exons do not overlap (Fig. S1D–F). To make a clear distinction between these two types of pairing arrangement, we term the former SAT and the latter bidirectional transcript (BDT). One of major differences between SAT and BDT is that SATs are capable of making double-stranded RNA between sense and antisense transcripts, while BDTs are not at least in the cytoplasm. The importance of the formation of double-stranded RNA between sense and antisense transcripts for stabilization of the sense transcript is reported in the recent reports [4,5,26]. Moreover, SATs are found less frequently in X chromosome than in autosome, but BDTs are found in similar proportion between autosome and X chromosome [7]. These observations infer some functional divergence between SATs and BDTs.

We performed *in silico* identification of SATs and BDTs (Fig. S1G) using the most recent murine transcript dataset [27]. Our first goal was to globally identify the expression modes of SATs, regardless of the association to promoter CpG island, by surveying the expression data obtained from 12 different normal mouse tissues (brain, thymus, heart, lung, liver, spleen, stomach, kidney, small intestine, testis and placenta [10.5 and 13.5 days postcoitum (dpc)]). Our previous study [8] has provided evidence of abundant nuclear transcripts for SATs and BDTs (presumably non-polyadenylated and non-spliced), so we monitored polyadenylated (poly(A)-positive) as well as total expressed transcripts (poly(A)-positive and poly(A)-negative) on a genome-wide microarray format for SAT expression analysis (covering 12,859 independent transcripts out of 4455 SATs and 2833 BDTs loci). We differentially labeled total RNA using oligo-dT (dT) versus random nanomer (Rd) primers. The dT method can detect mainly the poly(A)-positive population, whereas the Rd method can in principle detect almost all transcripts, including the poly(A)-negative population. The signals obtained by both the dT and Rd methods for each transcript showed good concordance, except for approximately one-third of transcripts (Fig. S2). Our microarray results obtained by both labeling methods were validated by confirming the microarray data using several different methods including real-time qPCR, Northern blot analysis and *in situ* hybridization [6,8,28].

If any relationships exist between sense and antisense transcripts then their expression among 12 tissues should show similar (simultaneous expression or silencing in a tissue, positive correlation) or reciprocal (exclusive expression in a tissue, negative correlation) patterns. However, if there is no relationship between sense and antisense transcripts then their expression patterns will be independent (no correlation). A considerable number of pairs show similar expression patterns (positive correlation) in both SATs and BDTs using dT-derived signals (Figs. 1A,C), consistent with recent studies [29–31].

SATs and BDTs can be categorized according to their relative orientation and degree of overlap; tail-to-tail (Fig. S1A,D), head-to-head (Fig. S1B,E) and embedded (Fig. S1C,F) [1]. The head-to-head population of dT signals showed the highest correlation, which may be due to sharing between the sense and antisense transcripts of part of the regulatory elements situated around the 5′ end of the first exon of each cDNA sequence (5′-FCS, which should in principle represent the putative TSS). This trend towards a positive correlation was also seen in tail-to-tail and embedded patterns, thus implying the existence of a *cis*-acting co-regulatory mechanism and/or a system that is working in a coordinated manner between sense and antisense transcription. In the case of Rd signals, although a predominant number of pairs showing a positive correlation, the distribution of correlation coefficients was not obviously different among the three mapping patterns (Figs. 1B,C). We also asked whether the observed expression modes were related to their protein-coding capacity, but there was no marked difference among them (Fig. S3). The existence of many positively correlated SATs and BDTs implies that many sense and antisense transcripts are not regulated independently but are regulated sufficiently to give simultaneous activation or inactivation of the gene pair.

All microarray data presented here can be browsed graphically via "Antisense Viewer" (http://www.brc.riken.go.jp/archives/Kiyosawa/Genomics_10/).

2.2. SATs and BDTs with an expression balance unique to a specific tissue are predominantly found in testis

Our previous studies demonstrated that for approximately 60% of SATs, the sense:antisense expression ratio fluctuated in different tissue and cell types in mice [6,8]. This differential expression implies that SATs and BDTs play a role in the maintenance of tissue-specific gene expression programs and/or possibly in cell differentiation. To search for SATs and BDTs that may be related to tissue specificity, we defined a "tissue-unique SAT or BDT (TU-SAT or TU-BDT)" as having a unique sense:antisense signal ratio among the tissues analyzed. We extracted