Ubiquitous antisense transcription in eukaryotes: novel regulatory mechanism or byproduct of opportunistic RNA polymerase?

Justin H Layer and P Anthony Weil*

Address: Department of Molecular Physiology and Biophysics, Vanderbilt University, School of Medicine, Nashville, TN 37232-0615, USA

* Corresponding author: P Anthony Weil (tony.weil@vanderbilt.edu)

The electronic version of this article is the complete one and can be found at: http://F1000.com/Reports/Biology/content/1/33
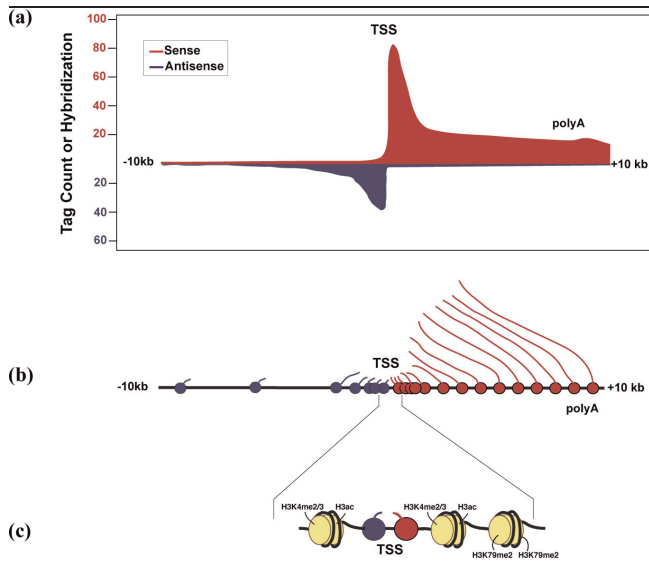
Abstract

One of the more notable observations made in the last few years in gene regulation is that eukaryotic genomes appear to be pervasively transcribed. Recent transcriptome mapping studies have shown that much of the genome is transcribed, and in some instances transcripts from both strands of specific genomic loci are detectable. While some of these transcripts map to known RNA polymerase II transcription units [that is, protein encoding open reading frames (ORFs)], many are derived from regions of DNA thought to be non-genic. Parallel chromatin immunoprecipitation studies of template-bound RNA polymerase II have shown that it is indeed resident on those regions found to be transcribed, both ORF and non-ORF. However, the strandedness of these pervasive transcripts has never been measured on a genome-wide basis. Four recent reports have addressed this question and, in the process, have made the startling discovery that many loci of mRNA sense gene transcription are associated with very active antisense or divergent transcription that begins at mapped transcription start sites and proceeds in an upstream direction.

Introduction and context

A truism in contemporary life-science research is that the questions that we, as experimentalists, can productively address are inextricably linked to the sophistication of our tools. As a corollary, when keen observation, creativity and new technology combine, exciting discoveries usually result. A striking example of these tenets is demonstrated by a series of papers published recently in *Science* [1–4]. Four groups have shown that the majority of transcribed mRNA encoding genes are also actively and divergently transcribed by RNA polymerease II (Pol II) in the direction opposite the open reading frame (ORF). These divergently transcribing ('antisense') polymerases may or may not be paused but, in contrast to the Pol II transcribing in the sense direction, they appear unable to elongate transcription of the antisense non protein-coding RNAs (ncRNAs) efficiently (Figure 1).

Current models hold that mRNA gene promoters are directional *cis*-active elements that drive transcription of linked downstream ORF-containing transcription units. The general Pol II transcription machinery, composed of general transcription factors (GTFs) such as Pol II, TFIIA, TFIIB, TFIID, TFIIE, TFIIF, and TFIIH, binds asymmetrically to the multiple elements that compose the promoter and transcription start site (TSS) [5]. GTFs and Pol II form a pre-initiation complex (PIC) on the promoter and initiation follows, with the RNA chain growing from the TSS. Under appropriate conditions the PIC then transitions through promoter escape and the ternary transcription complex elongates the nascent RNA chains until polyadenylation and termination occur. This simple model predicts: one directional promoter → one directional PIC → one directional transcript. The discovery of frequent Pol II promoter-proximal pausing [6–8] suggested a new level of regulatory complexity to this model of Pol II transcription. However, pausing *per se* does not explain the pervasive transcription in

Figure 1. Antisense transcription in human cells

(a) Data documenting antisense transcription. Shown is a summary of the data from the four reports [1–4] that this review focuses on, with particular emphasis on the data of Core *et al.* [1]. The red colored area designates transcription from the sense strand; the blue colored area represents antisense transcription and is quantified by negative values due to its antisense nature. On the ordinate is plotted either sequence count or hybridization signal depending upon the study. The more abundant sense transcripts (red) originate within a tight 50 bp window downstream of the typical mRNA-encoding gene TSS. TSS = 0, sequences downstream indicated to +10 kb, just downstream of the polyA site; sequences upstream of the TSS are indicated to −10 kb. Lower abundance antisense transcripts (blue) initiate across a broader window, and span a region of approximately 250 bp upstream of the sense TSS. The peaks of promoter-proximal transcripts represent paused, template-engaged Pol II molecules in the case of sense transcription; it remains to be tested if this is also true for antisense transcription. Abundant sense mRNA gene transcripts are detectable far downstream from the TSS, throughout the transcription unit, while antisense products are confined to a much more promoter-proximal region. **(b)** Molecular representation of sense and antisense transcription of mammalian mRNA-encoding genes. The data of **(a)** are depicted schematically with Pol II molecules densely localized (that is, 'paused') around the TSS. These are more abundant on the sense strand, and are shown in red; these Pol II molecules are also more densely localized about the TSS than their antisense counterparts shown in blue. These two types of transcription complexes may be biochemically distinct (red, blue; see text). Because the sense strand Pol II can transit promoter-proximal pause signals more efficiently, more template-engaged Pol II molecules are detected throughout the sense transcription unit, up to, and even past, the polyA site. **(c)** Correlation of antisense transcription with both active histone marks and Pol II occupancy. Three reports [1–3] identified a strong correlation between ChIP-measured Pol II occupancy and several histone epigenetic marks. There appears to be a nucleosome-free region between the peaks of sense and antisense TSSs. Biphasic Pol II ChIP signals peak just inside the apparent boundary nucleosomes shown. Notably, these boundary nucleosomes both carry 'active' histone epigenetic marks, such as H3ac and H3K4me2/3. However, only sense transcription is associated with histone epigenetic marks that correlate with Pol II transcription elongation (H3K79me2). bp, base pair; ChIP, chromatin immunoprecipitation; kb, kilobase; polyA, polyadenylation; Pol II; RNA polymerase II; TSS, transcription start site.

eukaryotes [9]. To begin to address this deficiency, the authors of four recent reports each developed or applied distinct techniques to monitor Pol II transcription on a global level [1–4]. Importantly, these techniques allowed for scoring transcription from both strands of DNA.

Major recent advances

Core *et al.* [1] devised a technique, termed global run-on sequencing, or GRO-seq, to isolate and sequence nascent transcripts. GRO-seq is an elegant adaptation of the classical nuclear run-on method. In this assay, chemically marked nascent RNA chains that had been labeled in isolated nuclei *in vitro* were purified and subsequently sequenced. The derived DNA sequence information was then used to elucidate the number, site, and direction of transcription of template-engaged polymerases across the human genome. Finally, this information was correlated with published Pol II and histone chromatin immunoprecipitation (ChIP) data. In primary human lung fibroblasts 16,882 NCBI Reference Sequence (RefSeq) protein coding genes were active (68% of annotated mRNA-encoding genes) and were scored, as predicted, as sense transcripts. Nearly all of these genes contained paused Pol II and exhibited promoter-proximal peaks of Pol II near the TSS. Amazingly though, over half of these transcription units also displayed antisense transcription (55% of all genes; 77% of active RefSeq genes). Divergent transcription (complexes) peaked just upstream of mapped RefSeq sense TSSs. Thus, antisense transcription is found at most actively transcribed mRNA encoding genes, which in humans also correlates very strongly with the presence of CpG islands. Further correlation of GRO-seq with Pol II ChIP and TSS data shows two peaks of Pol II that align with the sense and antisense peaks of nascent chains (Figure 1b), indicating that the majority of cognate transcription events are non-overlapping, and proceed from the RefSeq-mapped TSSs. Replotting the GRO-seq data with active, promoter-proximal histone modifications (histone H3Ac and H3K4me2) shows that both sense and antisense GRO-seq signals are bracketed by the modified histone ChIP signals, reinforcing the idea that GRO-seq does indeed score active transcription. Finally, transcription elongation histone marks (H3K36me3 and H3K79me3) are only associated with the ORF/body of sense mRNA-encoding genes (Figure 1c), consistent with the notion that sense-strand transcripts appear to be elongated much more efficiently than their divergent antisense counterparts.

Seila *et al.* [2] found short TSS-associated RNAs (TSSa-RNAs) that originate from regions surrounding the TSSs of known mRNA-encoding genes. They developed

methods to enrich and sequence TSSa-RNAs on a genome-wide basis and found that they mapped both downstream (within 50 nucleotides) and upstream (250 nucleotides) of TSSs. Many mRNA encoding genes in embryonic stem cells have both sense and antisense TSSa-RNAs. The frequency of production of TSSa-RNAs correlated with transcription levels. TSSa sense/antisense RNA ratios ranged from 1.5 to 2.5 for the most actively transcribed genes. The 20- to 90-nucleotide TSSa-RNAs co-localize with CpG islands, active histone marks and Pol II (ChIP signals). However, only the sense TSSa-RNAs correlate with the H3K79me2 transcription elongation chromatin mark, consistent with the idea that TSSa antisense transcripts are unable to elongate efficiently, as noted by Core *et al*. Finally, the authors show that the number of both sense and antisense TSSa-RNAs increase during repression of Oct4 levels in stem cells, suggesting dynamic regulation of this class of RNAs.

Preker and colleagues [3] depleted the RNA exosome in human cells using small interfering RNA knockdown and then examined gene-specific transcription via ENCODE (Encyclopedia of DNA Elements) tiling arrays. The rationale was to inhibit mRNA turnover to facilitate detection of putative low abundance, labile RNAs. As expected, exosome depletion increased known mRNA levels. However, a new class of short, polyadenylated, and highly unstable 500–2,500 nucleotide long promoter-proximal transcripts (PROMPTs) were detected. Subsequent analyses demonstrated that PROMPTs were transcribed from both sense and antisense strands extending from mapped TSSs of known RefSeq mRNA encoding genes. Experiments using synthetic reporter genes showed that PROMPT production requires a promoter (at least a 'sense' promoter; the cognate antisense promoter was not tested) and can be generated from heterologous DNA. These authors again correlated the production of PROMPT sense and antisense transcripts with (hypomethylated) CpG islands, and Pol II and H3K9ac ChIP data (ENCODE project). Interestingly, they also correlated PROMPT transcript levels with ChIP signals for E2F1 and TFIID {that is, the TFIID subunit TAF1 [TATA-binding-protein (TBP)-associated factor 1]} on a small subset of genes and found that these two factors tended to preferentially localize with PROMPT sense transcription. The authors interpreted these results (combined with their reporter gene data) to suggest that the generation of PROMPT antisense RNA synthesis may be independent of classic Pol II promoter elements/factors.

In the fourth paper, He *et al*. [4] describe their development of a technique termed asymmetric strand-specific analysis of gene expression (ASSAGE). This method, based on bisulfite-mediated deamination of RNA cytidine residues to uridine followed by conversion to cDNA and high throughput DNA sequencing, allows for the unambiguous assignment of the strandedness of the initial RNAs. Over 90% of useful sequencing reads were assignable to known RefSeq genes and were roughly equally split between sense and antisense orientations. These overall patterns were obtained in five human cell types analyzed. Variations in the patterns of antisense and sense transcripts were observed, suggesting that the antisense transcriptome is dynamic. Antisense transcripts (sequence tags) were preferentially derived from promoter (5′) and termination (3′) regions of ORFs, results consistent with the reports above.

Future directions
The work outlined above has clearly shown that the majority of mammalian mRNA encoding genes have template engaged, active Pol II molecules both up- and downstream of their major TSSs. Neither the role nor mechanism of formation of the template-engaged antisense paused Pol II molecules are known. Future experiments will dissect this process further by elucidating both the *cis*- and *trans*-active components responsible. Systematic high-resolution ChIP studies to score occupancy of GTFs, Pol II [including CTD (carboxy-terminal domain) isoforms] and other transcription factors across the TSS regions on genes *in vivo* will serve to identify many of the factors that may contribute to pausing, antisense transcription, and elongation competency. It is likely that the sense and antisense transcription complexes have biochemically distinct compositions. Genetic and biochemical experiments can also be used to address a few other questions. Which promoter element(s) direct the synthesis of antisense ncRNAs? Can the synthesis of ncRNAs be recapitulated using existing crude or reconstituted *in vitro* transcription systems? Similarly, genetic screens designed to identify contributing factors could be based upon a single promoter-TSS flanked by two readily scored reporter genes.

The real question here is whether antisense transcription plays a key role in gene regulation or, alternatively, as has been proposed previously [10], is simply representative of opportunistic Pol II-mediated transcription within nucleosome-free regions of the genome. Nucleosome-free regions are frequently found upstream of PICs/TSSs [11,12]. It should be noted that pervasive bidirectional transcription of nucleosome-free intergenic regions of DNA have been recently observed in the budding yeast *Saccharomyces cerevisiae* [13,14]. Perhaps the unique tools

available in this system will allow for inroads into the dissection of this intriguing process. Further, it appears that the abundant ncRNAs described above may well be actively modified by post-transcriptional processes, resulting in a dramatic new level of complexity of ncRNAs [15]. Thus, although the intricacies of eukaryotic gene regulation have increased exponentially through the discovery of ubiquitous antisense RNA synthesis, one thing is clear, many more exciting discoveries are ahead in the study of Pol II transcription!

Abbreviations

ASSAGE, asymmetric strand-specific analysis of gene expression; ChIP, chromatin immunoprecipitation; CTD, carboxy-terminal domain; ENCODE, Encyclopedia of DNA Elements; GRO-seq, global run-on sequencing; GTF, general transcription factor; ncRNA, non protein-coding RNA; ORF, open reading frame; PIC, pre-initiation complex; Pol II, RNA polymerase II; PROMPT, promoter-proximal transcript; RefSeq, NCBI Reference Sequence; TAF1, TATA-binding-protein associated factor 1; TBP, TATA binding protein; TSS, transcription start site; TSSa-RNA, TSS-associated RNA.

Competing interests

The authors declare that they have no competing interests.

References

1. Core LJ, Waterfall JJ, Lis JT: **Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters.** *Science* 2008, **322:**1845-8.

2. Seila AC, Calabrese JM, Levine SS, Yeo GW, Rahl PB, Flynn RA, Young RA, Sharp PA: **Divergent transcription from active promoters.** *Science* 2008, **322:**1849-51.

   F1000 Factor 6.0 *Must Read*
   Evaluated by Bino John 04 Feb 2009

3. Preker P, Nielsen J, Kammler S, Lykke-Andersen S, Christensen MS, Mapendano CK, Schierup MH, Jensen TH: **RNA exosome deple-tion reveals transcription upstream of active human promo-ters.** *Science* 2008, **322:**1851-4.

4. He Y, Vogelstein B, Velculescu VE, Papadopoulos N, Kinzler KW: **The antisense transcriptomes of human cells.** *Science* 2008, **322:**1855-7.

5. Juven-Gershon T, Hsu JY, Theisen JWM, Kadonaga JT: **The RNA polymerase II core promoter - the gateway to transcription.** *Curr Opin Cell Biol* 2008, **20:**253-9.

6. Zeitlinger J, Stark A, Kellis M, Hong JW, Nechaev S, Adelman K, Levine M, Young RA: **RNA polymerase stalling at develop-mental control genes in the *Drosophila melanogaster* embryo.** *Nat Genet* 2007, **39:**1512-6.

   F1000 Factor 6.5 *Must Read*
   Evaluated by Donald Rio 23 Nov 2007, Bing Ren 14 Dec 2007, Jeffrey Macklis 12 Feb 2008

7. Hendrix DA, Hong JW, Zeitlinger J, Rokhsar DS, Levine MS: **Promoter elements associated with RNA Pol II stalling in the *Drosophila* embryo.** *Proc Natl Acad Sci U S A* 2008, **105:**7762-7.

   F1000 Factor 3.0 *Recommended*
   Evaluated by Tony Weil 05 Jun 2008

8. Gilmour DS: **Promoter proximal pausing on genes in metazo-ans.** *Chromosoma* 2009, **118:**1-10.

9. Kapranov P, Willingham AT, Gingeras TR: **Genome-wide tran-scription and the implications for genomic organization.** *Nat Rev Genet* 2007, **8:**413-23.

10. Struhl K: **Transcriptional noise and the fidelity of initiation by RNA polymerase II.** *Nat Struct Mol Biol* 2007, **14:**103-5.

    F1000 Factor 6.0 *Must Read*
    Evaluated by Jonathan R Warner 09 Feb 2007

11. Mavrich TN, Jiang C, Ioshikhes IP, Li X, Venters BJ, Zanton SJ, Tomsho LP, Qi J, Glaser RL, Schuster SC, Gilmour DS, Albert I, Pugh BF: **Nucleosome organization in the *Drosophila* genome.** *Nature* 2008, **453:**358-62.

    F1000 Factor 6.4 *Must Read*
    Evaluated by Tony Weil 22 May 2008, Ian Dunham 05 Dec 2008

12. Schones DE, Cui K, Cuddapah S, Roh TY, Barski A, Wang Z, Wei G, Zhao K: **Dynamic regulation of nucleosome positioning in the human genome.** *Cell* 2008, **132:**887-98.

    F1000 Factor 9.0 *Exceptional*
    Evaluated by Paul Liu 29 Apr 2008

13. Xu Z, Wei W, Gagneur J, Perocchi F, Clauder-Münster S, Camblong J, Guffanti E, Stutz F, Huber W, Steinmetz LM: **Bidirectional promoters generate pervasive transcription in yeast.** *Nature* 2009, **457:**1033-7.

14. Neil H, Malabat C, d'Aubenton-Carafa Y, Xu Z, Steinmetz LM, Jacquier A: **Widespread bidirectional promoters are the major source of cryptic transcripts in yeast.** *Nature* 2009, **457:**1038-42.

15. Affymetrix ENCODE Transcriptome Project; Cold Spring Harbor Laboratory ENCODE Transcriptome Project: **Post-transcriptional processing generates a diversity of 5′-modified long and short RNAs.** *Nature* 2009, **457:**1028-32.