CellPress

Review

The Determinants of Directionality in Transcriptional Initiation

Dia N. Bagchi[1] and Vishwanath R. Iyer[1,*]

A new paradigm has emerged in recent years characterizing transcription initiation as a bidirectional process encompassing a larger proportion of the genome than previously thought. Past concepts of coding genes thinly scattered among a vast background of transcriptionally inert noncoding DNA have been abandoned. A richer picture has taken shape, integrating transcription of coding genes, enhancer RNAs (eRNAs), and various other noncoding transcriptional events. In this review we give an overview of recent studies detailing the mechanisms of RNA polymerase II (RNA Pol II)-based transcriptional initiation and discuss the ways in which transcriptional direction is established as well as its functional implications.

What is Bidirectional Transcription?

The determinants of transcriptional initiation are intricate and interwoven. What is clear from the high proportion of the human genome that is transcribed (estimated at 60%) compared with the small proportion that is coding (2%) [1,2] is that transcriptional processes involve much more of the genome than was once thought. High-resolution analyses and detailed catalogs of transcription start sites (TSS) obtained using next-generation sequencing methods have shown that transcription initiation frequently occurs in both directions from a given **promoter** (see Glossary) region [3,4]. These studies have raised the question of whether transcription initiation is an inherently **bidirectional** or unidirectional process. In one model, biases in the direction of transcription arise as emergent properties from the complex regulatory restrictions placed on inherently bidirectional promoter elements. In an alternative model, transcription at its core is unidirectional, with the appearance of bidirectionality arising due to the adjacent placement of individual unidirectional **core promoters** in opposite orientations. In the latter model, the similar needs of two separate gene promoters to coordinately regulate transcription factor (TF) recruitment might select for divergent transcript orientations. Transcription occurring in two directions from a single core promoter and **divergent transcription** originating from two distinct core promoters have not always been well distinguished in the literature. The conflation of these two categories has led to some ambiguity. Here we refer to transcription arising from a core promoter in opposite directions as bidirectional, whereas transcription of two outward-facing transcripts from independent core promoters is termed divergent transcription. To some extent, the terminology that researchers in the field adopt depends on variability in the definitions and size estimates of what constitutes a promoter and how far divergent genes may lie from one another. In this review we discuss the evidence for each model to illustrate the current understanding of transcriptional initiation and also consider the related issue of sense and antisense transcription at genes. Ultimately, we suggest a more nuanced view of promoters as non-directional, conducive regions of DNA prone to the occurrence of an open chromatin structure, the transcriptional potential of which is channeled either bidirectionally or unidirectionally in a

Trends

Most core promoters are bidirectionally competent with respect to initiation.

Promoters and enhancers share common features that are conducive to transcriptional initiation.

Transcriptional directionality is determined in an additive and hierarchical way.

Noncoding transcription can alter coding transcription levels both by spreading specific chromatin modifications and by producing functional RNAs.

[1]Department of Molecular Biosciences, Center for Systems and Synthetic Biology, Institute for Cellular and Molecular Biology, University of Texas at Austin, Austin, TX 78712, USA

*Correspondence: vishy@utexas.edu (V.R. Iyer).

 http://dx.doi.org/10.1016/j.tig.2016.03.005

context-dependent manner. The regulatory constraints of the various layers of regulation then work additively to produce specific transcriptional states (Figure 1, Key Figure).

Assaying the Transcriptome

High-resolution methods to define the transcriptome have revealed that transcription initiates not only in the expected location downstream of promoters but also within promoter regions upstream of coding sequences and bidirectionally at active enhancers. Both of these sources of noncoding transcription generally produce short, unstable RNAs that are rapidly degraded [5,6]. Transcription has also been observed to originate within transcript bodies [7] and from the 3′ ends of genes in antisense orientation [8] (Figure 2). Numerous techniques have been used to detect nascent transcripts (Table 1). Unstable transcripts can be identified when RNA degradation pathways are inhibited causing the persistence of unstable RNAs [9,10]. These experiments have been used to interrogate the genomic sites of transcriptional initiation and classify them broadly into three types based on their bidirectional potential: stable/stable, stable/unstable, and unstable/unstable. These categories reflect the functional directionality of a promoter but do not specify whether initiation actually occurs in both directions. For an in-depth account of the various noncoding transcripts that have been described and the techniques that have been used to identify them, see the review by Steinmetz *et al.* [11].

eRNAs

In mammalian cells, transcriptional activity at enhancers is widespread and dynamically regulated, generally producing unstable transcripts in both directions when actively functioning as an enhancer [12]. It should be noted, however, that most of the putative enhancers identified by chromatin profiling have not been experimentally validated as functional. A small subset of enhancers produce stable **long noncoding RNAs (lncRNAs)** as one of their transcript pairs. During changing conditions or cell states, **eRNA** production is the most rapid and salient transcriptional response, preceding even the transcription of TFs in response to the change [13]. The question of whether most eRNAs are functional remains open. Post-transcriptional knock down of a handful of eRNAs has revealed cases where they are necessary for enhancing transcription at interacting genes [14] and for promoter–enhancer loop formation [15]. However, there are also many instances where knock down of these eRNAs does not inhibit the function of the enhancer [16]. On average, however, eRNA transcription is a good predictor of enhancer activity [12]. For further discussion of the potential functions of eRNAs, see Li *et al.* [17].

Promoter Upstream Transcripts (PROMPTs)

Within promoter regions, noncoding RNAs termed **PROMPTs** (Figure 2) have been detected after depletion of components of the exosome, an RNA degradation complex [4]. Similar transcripts have been noted by other groups and have been termed bidirectional noncoding RNAs (BNCs) [18], cryptic unstable transcripts (CUTs) [3], or stable unannotated transcripts (SUTs) [3,11]. In mammalian cells, PROMPTs have been observed to be transcribed in both the sense and antisense directions [10,19]. These transcripts are generally transcribed by RNA Pol II but can also originate upstream from Pol I- and Pol III-transcribed genes [4]. Antisense PROMPT transcription has been reported to be correlated [10] and anticorrelated [4] with downstream coding genes. Skewing in initiation direction may reflect trade-offs where the presence of activated open chromatin generally recruits more of the transcription machinery, but also where a transcript's abundant expression may monopolize the pool of available RNA Pol II. In contrast to stable mRNAs, most PROMPTs and eRNAs are depleted for 5′ splice sites [5] and enriched for polyadenylation sites [6], features that target them for early transcriptional termination and degradation. While most PROMPTs are rapidly degraded, some stable noncoding transcripts produced from promoter regions have been shown to be functional [20,21]. Some promoter

Glossary

+1 nucleosome: the first nucleosome that appears within the DNA encoding a transcript, appearing immediately downstream of the gene's NDR.
−1 nucleosome: the first nucleosome that appears upstream of a gene's NDR.
Bidirectional transcription: transcription emanating outward from a single core promoter region in opposite directions.
Convergent transcription: transcription originating from opposite strands of DNA with the direction of transcription converging.
Core promoter: DNA sequence, generally found within a NDR, with the ability to independently initiate transcription.
Core promoter element: DNA sequence located within a promoter region that regulates transcriptional output.
Divergent transcription: transcription emanating outward from two separate core promoter regions in opposing directions.
Enhancer RNA (eRNA): RNA that is usually bidirectionally transcribed from intergenic regions that act as enhancers.
H2A.Z: histone variant that can replace the canonical H2A histone within nucleosomes and is associated with sites of transcriptional initiation.
Long noncoding RNAs (lncRNAs): stable noncoding RNAs that are produced either upstream of coding genes or from other intergenic initiation sites. Many have been shown to have structural or regulatory functions.
Nucleosome-depleted region (NDR): area of the genome associated with the presence of a TSS that is depleted for nucleosomes in the middle and typically has strongly positioned nucleosomes on either side.
Promoter: regulatory region of DNA occurring upstream of a transcript and comprising a core promoter and core promoter elements; required for the proper regulation of gene expression.
Promoter upstream transcripts (PROMPTs): noncoding transcripts originating from within a promoter region.