Pervasive transcription

A. Abhinaya

Pervasive transcription or interleaved transcription is the transcription of the interspersed genes which are embedded within the normal coding sequence. The quintessential factor is that, it is believed that the entire stretch of the genome is transcribed, whether it is coding for a particular protein or not. The other underlying factor is that not all coding sequences lie juxtaposed; they may also overlap one another. In other words, they may lie interspersed. The unconventional fact is that these overlapping sequences may be a non-coding region or an interspersed coding region which is still transcribed. The question is, how often are these interspersed regions getting transcribed and does all of the transcribed sequence get translated into protein coding messages?

Transcription is a term associated with the genic region. However, recent findings help extend this concept to include intergenic region, many non-functional elements, pseudogenes, etc. The shear size of the mRNA transcripts produced does not correspond to the mRNA translated; neither do the data relating to the number of coding regions correspond to the multitude of transcripts obtained. The genomic framework and transcription process have to be reviewed for better understanding of transcription of interleaved and intragenic regions. Non-coding, stably unannotated transcripts (SUTs) are also produced, which may have a vital function as a regulatory molecule. Some of these might translate to peptides and hence are called transcripts of unknown function (TUFs). Projects such as ‘ENCODE’ are aimed at exploring what is transcribed to what is translated to what is expressed. The results of the ENCODE project revealed new sectors of RNA and new layers of the transcription machinery. However, the precision of the positioning of the CUTs and SUTs is still to be studied, because of the diverse mechanism of generation of transcripts and also that the number of transcripts generated does not correlate with the findings of Xu et al. Some interesting cases that involve promoter-associated pervasiveness are discussed below, as they are the major source of SUTs.

Multiple transcription start site (TSS) points, which control transcription of single annotated genes, are found by mapping the 5’-end of capped RNAs to their corresponding TSS. TSS shifts result in the production of snRNAs and snoRNAs, as seen in the shift between upstream and downstream TSS of IMD2 gene in yeasts. Additionally, pervasive
transcription is also associated with the coding region15. Putative promoters found within retrotransposons (154 in mice and 579 in humans) are a source of novel isoforms of protein-coding genes15,19. Internal promoters found within intron-7 in mouse c-erbA gene or long interspersed nuclear-like retro-transposon called I factor produced a new transcript20. These are activated by the class-III group of RNA polymerase II (ref. 21). They help duplicate downstream genes which may not lie within the control of upstream promoters in operons or other genes, thereby transcribing CUTs which are exon-originated during the process22,23.

In case of bidirectional promoters, the interleaved sequence is associated with the nucleosome-free region and is often found associated with a promoter of another/adjacent protein-coding sequence, thereby promoting the activation of such regions24. PROMPTS are produced in this way18.19. A cryptic promoter also promotes transcription of a comparatively silenced region, promotion of open chromatin region by modulating the de-acetylases which modify the histone proteins25.

Apart from these varied promoters there are other factors which might influence transcription. In promoter occlusion of SER3 gene, the RNA polymerase II fails to terminate at the termination site of an upstream gene due to the presence of SRG1, and causes it to interfere with the promoter of SER3, but transcribes SRG1 (refs 1, 18). Stalled polymerases are also responsible for producing tRNAs generated from TSS of several protein-coding genes18,26. Genome-wide analysis reveals that nucleosome occupancy within the exon is higher compared to pseudogenes27. Therefore, modulating the nucleosomes will aid in exposing the exons for transcription. Additionally, pseudogenes, overlapping transcripts and intergenic regions also have a fair chance of getting transcribed. The curvature and bendability of the DNA may also aid in transcription initiation in this region28.

About 68% of these unannotated transcripts are associated with the 5′-nucleosome-free region aiding the transcription of sense PARs, but most of them are in the antisense direction13. Usually 70% of the human genes have at least one antisense transcript that overlaps the coding transcript. This leads to the formation of RNA–RNA hybrids, which induce RNAI-mediated silencing as in XIST29. RNA polymerase V imparts silencing of overlapping/adjacent genes by transcribing the intragenic region and the transcripts act as silencers30. Modification of the histones is also involved in repressing the overlapping genes29. The intragenic transcripts at the 3′-end are eliminated by the histone modification factors and histone chaperones (Spt6, Spt16 and HET set-2)1. But the most common form of eviction of these cryptic transcripts is via the exosome complex with the aid of the TRAMP complex31,32. Here too the complexity lies in which binding proteins (KRSP or TTP) bind to the ARE elements, based on which the transcript is either degraded or accumulated33.

These transcription procedures still account for only a meagre level. Future experiments may prove that the entire genome is being transcribed and that there might be more such layered mechanisms that meets the eye. With the onset of the minimal genome concept, it is more certain that these unannotated transcripts are involved in the regulatory affairs of the cell just like ncRNAs or SUTs, rather than being translated. It is important to study this aspect and explore the role of the non-coding transcripts as the expression of some ncRNAs is related to cancer. Possibilities of harbouring ncRNAs/SUTs that aid in avoiding illness may be a reality.