Intragenic antisense transcription correlates with long UTRs

Finocchiaro G (1,2), Parise P (1,2), Di Ninni V (1,2), Francois S (1,2),
Carro MS (1,2), Muller H (1,2)

(1) FIRC Institute of Molecular Oncology Foundation (IFOM), Milano
(2) European Institute of Oncology (IEO), Milano

Motivation

Studies on the distribution of transcription factor binding sites (TFBS) along entire chromosomes by chromatin immunoprecipitation have revealed that the majority of binding sites for several transcription factors lie far away from annotated promoters. Evidence is accumulating that the amount of transcribed DNA may be much higher than previously thought, with a significant fraction of transcription occurring in the antisense direction of annotated genes. Deciphering the regulatory potential of antisense transcripts is still in its infancy.

Methods

We performed genomic mapping of the 5'-ends of antisense transcripts to corresponding sense transcripts with the aim of identifying hotspots of intragenic antisense transcription. Correct orientation of antisense transcript was evaluated using criteria simlar to those adopted by [1]. When multiple antisense transcripts were mapped close to each other, they were assigned to distinct Antisense Transcription Starting Regions (ATSR) if their 5'ends were more than 500 bp apart. Otherwise they were classified as being part of the same ATSR. UTR length was estimated based on annotations available at :
http://hgdownload.cse.ucsc.edu/goldenPath/hg17/database/refSeqSummary.txt.gz,
DBTSS (http://dbtss.hgc.jp/), and a recently published curated dataset of 5'UTRs [2], with highly consistent results.

Results

A total of 7903 ATSRs was identified that were mapped onto 5075 genes. In agreement with published results, we find that antisense transcripts are particularly abundant at the 5' end as well as at the 3' end of genes. However, our map identifies the first exon, the 5'end of the first intron, and the 5'end of the last exon as hotspots of intragenic antisense transcription when the number of antisense transcripts in a given region is normalized per unit sequence. In particular, we identified 654 genes with an ATSR in their first exon, 1847 genes with an ATSR in their first intron, and 756 genes with an ATSR in their last exon. The remaining ATSRs are distributed evenly along loci. Our findings are supported by the enrichment of known transcription factor binding sites in the vicinity of ATSRs as well as by binding of the general transcription factor TAF1 as measured by chromatin immunoprecipitation on genomic tiling arrays [3]. The vicinity of hotspots of antisense transcription to the UTRs of sense transcripts prompted us to explore the UTRs of genes with evidence of intragenic antisense transcription in more detail. We find that the presence of antisense transcripts is strongly and positively correlated with the length of UTRs of the corresponding sense transcript. While the median length of 5'UTRs of genes in the genome was found to be 142 bp, genes with an ATSR in their first exon had a median 5'UTR of 241 bp. A T-test performed on the log transformed values of 5'UTR lengths indicates that this is significant at P = 7.67E-24. Similarly, while the median length of 3'UTRs in the genome was estimated to be 636 bp, genes with an ATSR in their last exon had a 3'UTR of median length 1225 bp. T-test on log transformed 3'UTR lengths indicates that this result is significant at P = 1.63E-29. Potential implications of this finding for the regulatory role of antisense transcripts will be discussed.

Contact email: heiko.muller@ifom-ieo-campus.it

References

1. Chen J, Sun M, Kent WJ, Huang X, Xie H, Wang W, Zhou G, Shi RZ, Rowley JD: Over 20% of human transcripts might form sense-antisense pairs. Nucleic Acids Res 2004, 32(16):4812-4820.
2. Davuluri RV, Suzuki Y, Sugano S, Zhang MQ: CART classification of Human 5'UTR Sequences. Genome Research 2006, 10:1807-1816.
3. Kim TH, Barrera LO, Zheng M, Qu C, Singer MA, Richmond TA, Wu Y, Green RD, Ren B: A high-resolution map of active promoters in the human genome. Nature 2005, 436(7052):876-880.