Bioinformatics, 35(15), 2019, 2529–2534
doi: 10.1093/bioinformatics/bty981
Advance Access Publication Date: 10 December 2018
Discovery Note

OXFORD

Genome analysis

Deciphering the complexity of human non-coding promoter-proximal transcriptome

Sarah N. Mapelli[1,2,3], Sara Napoli[1], Giuseppina Pisignano[1], Ramon Garcia-Escudero[1,4], Giuseppina M. Carbone[1] and Carlo V. Catapano[1,3,5,*]

[1]Tumor Biology and Experimental Therapeutics Program, Institute of Oncology Research (IOR), Università della Svizzera Italiana (USI), Bellinzona 6500, Switzerland, [2]Institute of Computational Science, Università della Svizzera Italiana (USI), Lugano 6900, Switzerland, [3]Swiss Institute of Bioinformatics (SIB), Lausanne 1015, Switzerland, [4]Molecular Oncology Unit-CIEMAT and Centro de Investigacion Biomedica en Red de Cancer (CIBERONC), Madrid 28040, Spain and [5]Department of Oncology, Faculty of Biology and Medicine, University of Lausanne, Lausanne1066, Switzerland

*To whom correspondence should be addressed.

Associate Editor: John Hancock

Received on July 3, 2018; revised on November 8, 2018; editorial decision on November 26, 2018; accepted on December 7, 2018

Abstract

Motivation: Long non-coding RNAs (lncRNAs) have gained increasing relevance in epigenetic regulation and nuclear functional organization. High-throughput sequencing approaches have revealed frequent non-coding transcription in promoter-proximal regions. However, a comprehensive catalogue of promoter-associated RNAs (paRNAs) and an analysis of the possible interactions with neighboring genes and genomic regulatory elements are missing.

Results: Integrating data from multiple cell types and experimental platforms we identified thousands of paRNAs in the human genome. paRNAs are transcribed in both sense and antisense orientation, are mostly non-polyadenylated and retained in the cell nucleus. Transcriptional regulators, epigenetic effectors and activating chromatin marks are enriched in paRNA-positive promoters. Furthermore, paRNA-positive promoters exhibit chromatin signatures of both active promoters and enhancers. Promoters with paRNAs reside preferentially at chromatin loop boundaries, suggesting an involvement in anchor site recognition and chromatin looping. Importantly, these features are independent of the transcriptional state of neighboring genes. Thus, paRNAs may act as *cis*-regulatory modules with an impact on local recruitment of transcription factors, epigenetic state and chromatin loop organization. This study provides a comprehensive analysis of the promoter-proximal transcriptome and offers novel insights into the roles of paRNAs in epigenetic processes and human diseases.

Availability and implementation: Genomic coordinates of predicted paRNAs are available at https://figshare.com: https://doi.org/10.6084/m9.figshare.7392791.v1 and https://doi.org/10.6084/m9.figshare.4856630.v2.

Contact: carlo.catapano@ior.usi.ch

Supplementary information: Supplementary data are available at *Bioinformatics* online.

1 Introduction

The human genome generates thousands of non-coding transcripts (Mercer and Mattick, 2013). Despite the initial concern that most of the non-coding transcriptome would lack specific functions, evidence is rapidly emerging that many long non-coding RNAs (lncRNAs) play important biological functions and take part in

genomic reprogramming during development and diseases (Mercer and Mattick, 2013). High-throughput studies using tiling arrays and RNA sequencing (RNA-Seq) have revealed the presence of non-coding transcripts in promoter-proximal regions (Core *et al.*, 2008; Kapranov *et al.*, 2007; Preker *et al.*, 2008; Seila *et al.*, 2008). Indeed, a large fraction of non-coding transcription occurs within the region immediately upstream the transcription start site (TSS) of genes (Flynn *et al.*, 2011; Sigova *et al.*, 2013, 2015). Because of low transcription rates and rapid degradation, promoter-proximal RNAs have been difficult to detect (Flynn *et al.*, 2011; Preker *et al.*, 2008). Furthermore, a genome-wide catalogue of promoter-associated RNAs (paRNAs) along with a comprehensive analysis of the potential interactions with neighboring genes and genomic regulatory elements is missing (Guil and Esteller, 2012). Providing a large catalogue of annotated paRNAs and of their putative interacting partners, our study can set the framework for investigating the functional and structural contribution of paRNAs in genetic and epigenetic mechanisms of human diseases.

2 Results

2.1 Definition of the promoter-proximal transcriptome in the human genome

To search for novel transcripts in promoter-proximal regions we built a compendium of global nuclear run-on sequencing (GRO-Seq) data from human cell lines (Allen *et al.*, 2014; Li *et al.*, 2013; Sigova *et al.*, 2013; Yang *et al.*, 2013). Then, we applied bioinformatics tools for transcript identification (Heinz *et al.*, 2010) and searched for novel transcripts within the 2-kb promoter-proximal regions of human genes (Fig. 1A).

We limited the analysis to annotated genes with an intergenic distance of ≥10 kb on both strands to avoid mapping transcripts from overlapping transcriptional units. About 14 000 promoters in the Hyperbrowser database (https://hyperbrowser.uio.no/hb/) fulfilled these criteria. Using this approach, we identified thousands of promoter-proximal transcripts in the five initial datasets analyzed (Fig. 1B). To assess the reproducibility of these findings, we processed 41 additional GRO-Seq datasets from published studies (Supplementary Tables S1 and S2) increasing the catalogue to 34 cell lines and 7 distinct treatment conditions. In line with previous studies (Core *et al.*, 2008; Kapranov *et al.*, 2007; Preker *et al.*, 2008; Seila *et al.*, 2008), we found that AS transcripts outnumbered S transcripts (Fig. 1B). This finding was supported analysis of the additional cell lines in the validation set (Supplementary Fig. S1A). Indeed, many promoters had only AS paRNAs (Fig. 1C and Supplementary Fig. S1B). However, we also found a substantial number of promoters that had exclusively or concomitantly S paRNAs. AS paRNAs mapped closer to the gene TSS with most transcripts located at ≤0.5 kb, whereas S paRNAs spread more broadly within the promoter region with many transcripts ≥1 kb from the gene TSS (Fig. 1D–E). Similar distributions of S and AS paRNAs were observed in multiple cell lines (Supplementary Figs S2 and S3).

About one third of the interrogated 14 000 promoters (n = 5009) had either S or AS transcripts in at least one cell line (Fig. 1F). About a quarter of promoters had exclusively AS paRNAs whereas one tenth of the interrogated promoters had exclusively or concomitantly S paRNAs. A fraction of paRNA-positive promoters (r = 0.14) shared similar transcripts among multiple cell lines (Fig. 1F). The fraction of shared transcripts was higher for AS (r = 0.06) than S (r = 0.01) paRNAs. However, most paRNAs had a more restricted expression pattern. Interestingly, genes with shared paRNAs, irrespective of their strand orientation, had lower coefficient of

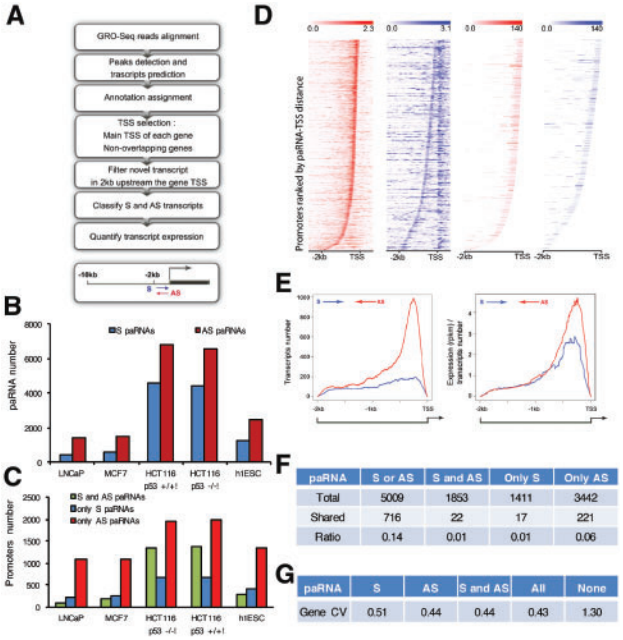

Fig. 1. Promoter-associated transcriptome in the human genome. (**A**) Workflow for the identification and classification of promoter-proximal transcripts from GRO-Seq data. (**B**) Number of predicted S and AS paRNAs in individual cell lines. (**C**) Number of promoters expressing S paRNAs, AS paRNAs or both S and AS paRNAs in individual cell lines. (**D**) Density plots of GRO-Seq reads (*left panels*) and predicted promoter-proximal transcripts (*right panels*) in h1ESCs ordered by distance from the adjacent gene TSS. Left panels, AS transcripts; Right panels, S transcripts. (**E**) Cumulative distribution of promoter-proximal transcripts based on their number (*left panel*) and normalized expression measure as RKPM/total number of transcripts (*right panel*) in h1ESCs. (**F**) Number and ratio of promoters with total or shared (present in all cell lines, as in panel C) S and/or AS paRNAs. (**G**) Coefficients of variation (CV) of the expression of genes with without or with shared paRNAs across all cell lines

variation (CV) (CV = 0.44) compared to genes without paRNA (CV = 1.30), suggesting paRNAs marked genes with reduced expression fluctuation across cell lines. (Fig. 1G). Increasing the number of cell types analyzed the number of promoters (n = 6575) exhibiting paRNA in at least one dataset increased. Conversely, the fraction of promoters with shared paRNA in all cell lines decreased.

2.2 Promoter-proximal transcripts and their relation with adjacent genes

GRO-Seq detected discrete transcripts with AS and S orientation in individual promoters (Fig. 2A). In many cases, we found an overlap of the predicted transcripts across multiple cell lines suggesting a certain degree of conservation (Fig. 2B; Supplementary Figs S4 and S5). However, individual promoters could express either S or AS paRNAs in different cell lines, underlying the dynamic and cell context-specific nature of the transcripts. To verify our predictions, we performed quantitative reverse transcriptase-polymerase chain reaction (qRT-PCR) and directional strand-specific RT-PCR (ssRT-PCR) using primers spanning the promoter-proximal regions of selected genes (n = 16). paRNAs were detected at all the predicted positions ($n = 26$), confirming the validity of our search strategy (Supplementary Fig. S6A). The level of paRNA expression varied among the cell lines. Moreover, using directional ssRT-PCR, we found that transcript orientation was promoter- and cell type-specific (Supplementary Fig. S6B). Consistently, we recently

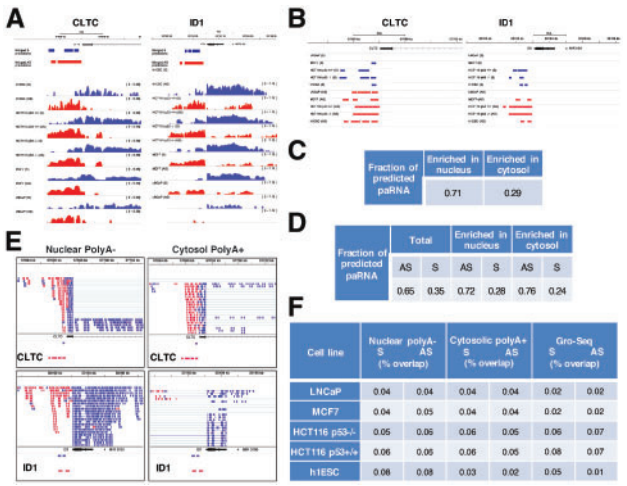

Fig. 2. Detection of promoter-associated transcripts by GRO-Seq and RNA-Seq. (**A**) GRO-Seq traces of nascent AS and S transcripts in CLTC and ID1 promoter. (**B**) GRO-Seq predicted promoter-associated transcripts in CLTC and ID1 gene in individual cell lines. Top, RefSeq gene annotation. (**C**) Fraction of GRO-Seq predicted paRNAs detected in nuclear polyA- and cytosolic polyA+ RNA-Seq data. (**D**) Fraction of GRO-Seq predicted S or AS paRNA detected by RNA-Seq in total, nuclear and cytosolic RNA. (**E**) Partitioning of CLTC and ID1 promoter-associated transcripts in nuclear polyA- and cytosolic polyA+ RNA from h1ESCs determined by RNA-Seq. (**F**) Fraction of promoter-associated transcripts overlapping adjacent gene transcripts in nuclear, cytosolic and nascent RNA

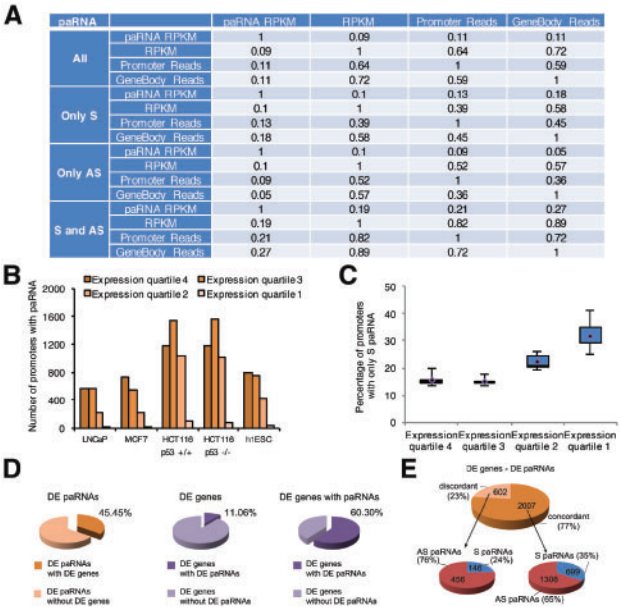

Fig. 3. Interdependence of promoter-associated RNAs and gene expression. (**A**) Pearson correlation values of expression of paRNAs, adjacent genes or gene sections in h1ESCs. Promoters, reads from -100 to +400 bp from the TSS; gene body, read from +1000 bp to the 3′-end of the gene. (**B**) Number of promoters with paRNAs in relation to the expression levels of the neighboring gene. (**C**) Percentage of promoters with S paRNAs in relation to the adjacent gene expression level. (**D**) Proportion of differentially expressed (DE) paRNAs (*left panel*), total DE genes (*middle panel*) and DE genes with paRNAs (*right panel*) with or without concomitant changes in expression in hESC1. (**E**) Proportion of concordant or discordant DE genes and paRNAs (*top*) and fractions of AS and S paRNAs (*bottom*) in hESC1

reported that the expression of S and AS paRNAs in the CDH1 promoter varied in different epithelial cell types in relation to the transcriptional state of gene (Pisignano *et al.*, 2017).

For assessing the functional relationship between paRNAs and adjacent genes we designed small interfering RNAs (siRNAs) targeting the predicted transcripts in three genes (CLTC, FERMT2 and HECTD1) in which we had detected the presence of paRNAs. SiRNA-mediated targeting of paRNAs can result in gene activation or repression depending on the promoter context, cell type and the gene transcriptional state (Kalantari *et al.*, 2016). Consistently, we showed recently activation of CDH1 transcription targeting of a regulatory paRNA with strand-specific siRNAs (Pisignano *et al.*, 2017). Using the same approach, we found that targeting the S transcripts in the CLTC promoter increased CLTC transcription in DU145 and PC3 cells (Supplementary Fig. S6C). Targeting the AS paRNAs with siRNAs reduced (FERMT2) or increased (HECTD1) the neighboring gene transcription. Thus, paRNAs can influence the adjacent genes likely through distinct mechanisms that depend on the promoter and cell context.

To further validate our predictions we took advantage of stranded RNA-Seq data of cytosolic and nuclear RNA from the ENCODE project (http://genome.crg.es/encode). The predicted paRNAs had deeper coverage (0.71 versus 0.29) in nuclear polyA$^-$ RNA than in cytosolic polyA$^+$ RNA (Fig. 2C). Furthermore, AS transcripts prevailed in all the fractions compared to S paRNAs (Fig. 2D). Inspection of individual promoters confirmed the overall prevalence of paRNAs in the nuclear polyA$^-$ RNA fraction (Fig. 2E; Supplementary Fig. S8). Thus, most of the predicted paRNAs were non-polyadenylated and retained in nuclei, in agreement with our recent analysis of CDH1-specific paRNAs (Pisignano *et al.*, 2017). Nevertheless, our analysis of RNA-Seq data shows also examples of promoter-proximal transcripts exported to the cytosol, hinting to possibly different functions.

Next, we assessed the possibility of overlap or continuity between paRNAs and gene transcripts. The presence of overlapping reads spanning paRNAs and S or AS transcripts in the gene's body would indicate that the predicted paRNAs are non-independent transcripts but extensions of gene transcripts. Inspection of individual genes did not reveal overlaps between paRNAs and gene transcripts (Fig. 2E; Supplementary Fig. S8). Genome-wide analysis of all promoters with predicted paRNAs showed that a very small fraction of paRNAs (0.02–0.08%) had evidence of continuity with gene transcripts (Fig. 2F). A similar analysis using GRO-Seq data gave comparable results. Thus, most paRNAs represent distinct transcripts unrelated to transcripts emanating from the adjacent genes.

2.3 paRNAs and neighboring genes are interconnected functional units

Evidence in support of a functional link between paRNAs and neighboring genes are gradually emerging (Sigova *et al.*, 2015). The relationship is likely complex involving multiple potential modes of interaction. We did not find a correlation between paRNA and gene expression (Supplementary Fig. S9). Pearson correlation coefficients were low for all the conditions tested in both the explorative and validating datasets, even for the promoter region (−0.1 to +0.4 kb) and gene body (+1 to +5 kb) (Fig. 3A; Supplementary Figs S10 and S11; Supplementary Table S3).

To better explore the relation between paRNAs and genes, we divided the genes in quartiles based on their expression level. This analysis showed that highly transcribed genes (quartile 3–4) were associated more frequently with paRNA-positive promoters (Fig. 3B; Supplementary Fig. S12). Highly transcribed genes

exhibit also a prevalence of promoters with AS paRNAs (Supplementary Table S4). Conversely, low transcribed genes (quartile 1) had a higher percentage of promoters with S paRNAs (Fig. 3C and Supplementary Fig. S13). These data strongly suggested a link between paRNAs and transcriptional state of the neighboring genes.

We examined further this aspect by interrogating GRO-Seq data from cells exposed to agents that could modulate the cell transcriptome. If paRNAs and genes were functionally connected, one would expect concomitant changes in expression of paRNA and gene pairs. In h1ESCs exposed to a transcriptional inducer (Sigova *et al.*, 2013), we found that 45% of genes with differentially expressed (DE) paRNAs exhibit changes in expression (Fig. 3D, *left panel*). On the other hand, considering DE gene exhibiting paRNAs 60% of them had changes in the corresponding paRNAs (Fig. 3D, *right panel*). Thus, in many cases paRNAs and genes responded concomitantly. Conversely, considering DE genes independently of the presence of paRNA, only 11% of them had paRNA expression changes (Fig. 3D, *middle panel*). The changes in DE paRNA-gene pairs were mostly concordant (increased or decreased concomitantly), whereas they were discordant in about a quarter of cases (Fig. 3E). Concordant and discordant pairs had similar distribution of S and AS paRNAs, indicating that the transcript orientation did not influence the changes. Analysis of gene and paRNA expression data from two additional cell lines (VCaP and K562) subjected to distinct treatments gave consistent results with many gene and paRNA pairs showing concomitant changes in expression independently of paRNA strand orientation (Supplementary Fig. S14). Together, these findings show the highly dynamic nature of the paRNA and gene interactions. Furthermore, the occurrence of both concordant and discordant changes in gene-paRNA expression indicates the possibility of both positive and negative interactions of S and AS paRNAs with the neighboring genes.

2.4 paRNAs act within a complex framework of transcriptional and epigenetic regulators

paRNAs likely act within complex regulatory circuits in a cell context-specific manner. In line with this, expressed genes (≥0.1 rpkm) with promoters with or without paRNAs were associated with distinct gene ontology pathways reproducibly in multiple cell lines (Supplementary Tables S5 and S6), suggesting a relation between paRNA distribution in the genome and cell-specific transcriptional programs. To explore the basis of paRNA-gene functional interactions, we assessed binding of transcriptional regulators to gene promoters using the ENCODE ChIP-Seq database. Positive transcriptional regulators (e.g. PHF8) were enriched in promoters with paRNAs (Fig. 4A; Supplementary Tables S7–S12), whereas repressive factors, like EZH2, were increased in promoters without paRNAs (Fig. 4B). Consistently, active chromatin marks (H3K27ac, H3K9ac, H3K4me3, H3K36me, H3K4me1) were enriched in promoters with paRNAs, whereas repressive marks (H3K27me3, H3K9me3) were not (Fig. 4C; Supplementary Table S13). Cumulative plots showed consistent differences in distribution as function of presence or absence of paRNAs (Fig. 4D–G; Supplementary Fig. S15). Interestingly, we observed a bimodal distribution of activating factors and histone marks with an upstream peak overlapping the putative position of paRNAs (Fig. 4D and E). This occurred similarly in presence of AS, S or both AS and S paRNAs. Conversely, EZH2 and the repressive histone marks

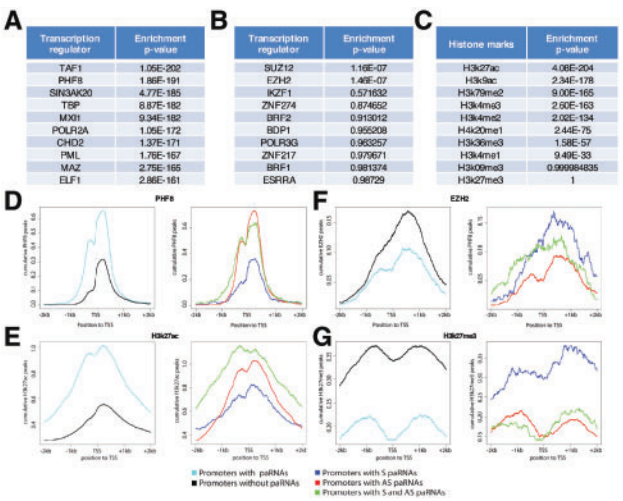

Fig. 4. Promoter-associated RNAs and distribution of transcriptional regulators and chromatin marks. (**A–B**) Transcriptional regulators (ranked by *P*-value based on hypergeometric test) enriched in promoters expressing (A) or non-expressing (B) paRNAs. (**C**) Histone marks (ranked by *P*-value) enriched in promoters expressing paRNAs. (**D**) Binding of PHF8 to gene promoters in relation to the presence of paRNAs in h1ESC. (**E**) Distribution of H3K27Ac in relation to the presence of paRNAs in h1ESC. (**F**) Binding of EZH2 in relation to the presence of paRNAs in h1ESC. (**G**) Distribution of H3K27me3 in relation to the presence of paRNAs in h1ESC

occupied prevalently promoters without paRNAs or with S paRNAs (Fig. 4F and G).

Slight differences in gene expression could affect our analysis (Supplementary Fig. S16A). To rule this out, we performed the same analysis in gene subsets equivalent in terms of gene numbers and expression levels (Supplementary Fig. S16B). Using these gene sets, we confirmed the differential binding of activating proteins and histone marks to paRNA-positive promoters (Supplementary Fig. S16C and D). Furthermore, we obtained consistent results in another cell line of the validation set (K562) for which we had available matching ChIP-Seq data (Supplementary Figs S17 and S18; Supplementary Table S13). Thus, binding of positive regulators occurred at promoters with both S and AS paRNAs and was independent of transcriptional state of the neighboring genes.

2.5 paRNAs are associated with promoter/enhancer local chromatin states and loop boundaries

We used combinatorial analysis of multiple epigenetic features (Ernst and Kellis, 2010) to explore the relationship between distinct chromatin states in gene promoters and paRNAs. Promoters with paRNAs had a broader or bimodal distribution of active TSS and promoter states (Fig. 5A). We found also enrichment of chromatin marks associated with active enhancers in paRNA-positive promoters (Fig. 5B). These features were similarly enriched in presence of AS and S paRNAs and thus were independent of the transcript orientation (Supplementary Fig. S18). Similar trends were observed examining paRNAs and matching data on chromatin marks in K562 cells (Supplementary Fig. S19A and B). An enhancer-like state of promoters expressing paRNA is consistent with previous studies pinpointing a role of divergent non-coding transcripts in *cis*-regulation of neighboring genes (Dao *et al.*, 2017; Engreitz *et al.*, 2016; Lepoivre *et al.*, 2013; Luo *et al.*, 2016). Indeed, our data extend this concept to both S and AS paRNAs with putative promoting and repressive capacity on the adjacent genes, leading to the hypothesis

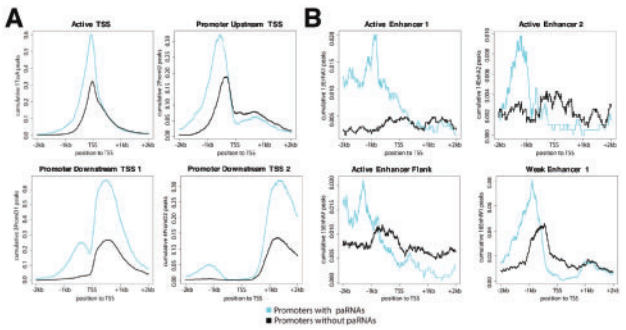

Fig. 5. Distinct epigenomic states and promoter-associated RNAs. (**A**) Distribution of promoter and TSS chromatin states in promoters with or without paRNAs in h1ESC. (**B**) Distribution of enhancer chromatin states in promoters with or without paRNAs in h1ESC

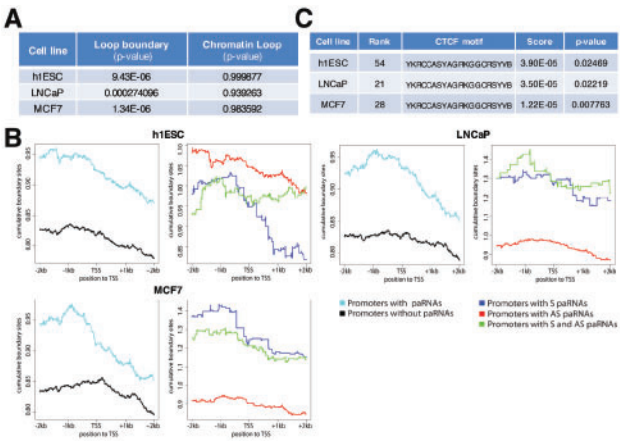

Fig. 6. Promoter-associated RNAs and distribution of chromatin loop anchor sites. (**A**) Enrichment of promoters with paRNAs (based on hypergeometric test, *P*-value) at loop boundary sites or chromatin loops. (**B**) Distribution of loop boundary regions in relation to the presence of paRNAs. (**C**) Directionality of enriched CTCF motifs in promoters with S paRNAs

that paRNAs might finely tune gene expression also through changes in chromatin architecture.

Promoters and enhancers play key roles in chromatin looping (Bonev and Cavalli, 2016). To explore the possibility that paRNAs might contribute to chromatin loop formation, we examined promoter partitioning in chromatin loops and boundaries interrogating Hi-C data (Rao *et al.*, 2014). Interestingly, we found enrichment of paRNA-positive promoters at loop boundaries but not within loops in two distinct cell lines (Fig. 6A, Supplementary Fig. S21A). The major interaction area was 2-kb upstream the TSS, overlapping the putative position of paRNAs (Fig. 6B). Importantly, we confirmed enrichment of paRNA-positive promoters at loop boundaries (*P*-value = 0.0048) by probing gene sets with equivalent expression levels (Supplementary Fig. S21B). Furthermore, average expression of paRNA-positive genes associated or not with loop boundaries was similar (Supplementary Fig. S21C), suggesting that their association with loop anchor sites did not depend on the gene transcriptional state.

We examined next the hypothesis that paRNAs might be implicated in recognition and binding of structural proteins, such as CTCF, at loop boundaries (Bonev and Cavalli, 2016). Interestingly, CTCF binding motifs are asymmetrical and pairs of convergent motifs at the anchor points are thought to drive loop formation (Rao *et al.*, 2014). We found a striking enrichment of CTCF binding motifs with S orientation in promoters with S paRNAs (Fig. 6C). Thus, paRNAs could favor anchor site recognition and loop formation by acting as scaffolds for architectural proteins and establishing interactions between distant genomic regulatory elements.

3 Discussion

In this study, we used an integrative bioinformatics approach to search for promoter-proximal transcripts in the human genome applying novel transcripts prediction tools to a large compendium of GRO-Seq datasets from more than 30 distinct cell types (Supplementary Fig. S22). We uncovered a large network of paRNAs that might complement the action of transcription factors and epigenetic effectors and participate in functional organization of the genome. Our data show that promoter-proximal transcription is pervasive in the human genome and generates a large repertoire of non-coding transcripts with potential regulatory functions. We found that more than a third of the 14 000 interrogated promoters had evidence of S or AS transcripts within the 2-kb space upstream the TSS of adjacent genes. Our analysis revealed a more complex and dynamic picture of the promoter-proximal transcriptome than shown by previous high-throughput studies on single or limited numbers of cell types. Most promoters exhibited multiple and often overlapping AS and S transcripts. The diversity and complexity of the possible interactions need to be taken in consideration when designing experiments to investigate individual paRNAs. We confirmed this picture by assessing paRNAs within a selected group of promoters in human cell lines and investigating the effects of siRNA-mediated paRNA targeting. The presence and strand orientation of paRNAs were not a fixed feature but changed dynamically depending on the cell type and experimental context and in relation to the transcriptional state of the neighboring genes. Furthermore, our analyses indicate that paRNAs are associated with a large number of genes and might function at multiple levels. Further studies will need to determine the molecular underpinnings of these complex interactions. Our present study provides a comprehensive and useful framework for investigating the role of paRNAs in diverse processes, ranging from promoter regulation and transcriptional control to chromatin state and organization.

Understanding the properties and the mechanisms of paRNA-based regulatory circuits may provide novel insights in fundamental aspects of transcriptional regulation and genome organization. Transcriptional and epigenetic mechanisms have a central role in many human diseases (Portela and Esteller, 2010). Assessing the promoter-proximal transcriptome and understanding their contribution to gene regulation and chromatin function might provide insights in the pathogenesis of human diseases and open to new therapeutic strategies.

Funding

This work was supported by Swiss National Science Foundation (Grant 310030-153099), Ticino Foundation for Cancer Research and Fidinam/Ceresio Foundation.

Conflict of Interest: none declared.

References

Allen,M.A. *et al.* (2014) Global analysis of p53-regulated transcription identifies its direct targets and unexpected regulatory mechanisms. *Elife*, **3**, e02200.

Bonev,B. and Cavalli,G. (2016) Organization and function of the 3D genome. *Nat. Rev. Genet.*, **17**, 772.

Core,L.J. *et al.* (2008) Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. *Science*, **322**, 1845–1848.

Dao,L.T.M. *et al.* (2017) Genome-wide characterization of mammalian promoters with distal enhancer functions. *Nat. Genet.*, **49**, 1073–1081.

Engreitz,J.M. *et al.* (2016) Local regulation of gene expression by lncRNA promoters, transcription and splicing. *Nature*, **539**, 452–455.

Ernst,J. and Kellis,M. (2010) Discovery and characterization of chromatin states for systematic annotation of the human genome. *Nat. Biotechnol.*, **28**, 817–825.

Flynn,R.A. *et al.* (2011) Antisense RNA polymerase II divergent transcripts are P-TEFb dependent and substrates for the RNA exosome. *Proc. Natl. Acad. Sci. USA*, **108**, 10460–10465.

Guil,S. and Esteller,M. (2012) Cis-acting noncoding RNAs: friends and foes. *Nat. Struct. Mol. Biol.*, **19**, 1068–1075.

Heinz,S. *et al.* (2010) Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol. Cell*, **38**, 576–589.

Kalantari,R. *et al.* (2016) Regulation of mammalian transcription and splicing by Nuclear RNAi. *Nucleic Acids Res.*, **44**, 524–537.

Kapranov,P. *et al.* (2007) RNA maps reveal new RNA classes and a possible function for pervasive transcription. *Science*, **316**, 1484–1488.

Lepoivre,C. *et al.* (2013) Divergent transcription is associated with promoters of transcriptional regulators. *BMC Genomics*, **14**, 914.

Li,W. *et al.* (2013) Functional roles of enhancer RNAs for oestrogen-dependent transcriptional activation. *Nature*, **498**, 516–520.

Luo,S. *et al.* (2016) Divergent lncRNAs regulate gene expression and lineage differentiation in pluripotent cells. *Cell Stem Cell*, **18**, 637–652.

Mercer,T.R. and Mattick,J.S. (2013) Structure and function of long noncoding RNAs in epigenetic regulation. *Nat. Struct. Mol. Biol.*, **20**, 300–307.

Pisignano,G. *et al.* (2017) A promoter-proximal transcript targeted by genetic polymorphism controls E-cadherin silencing in human cancers. *Nat. Commun.*, **8**, 15622.

Portela,A. and Esteller,M. (2010) Epigenetic modifications and human disease. *Nat. Biotechnol.*, **28**, 1057–1068.

Preker,P. *et al.* (2008) RNA exosome depletion reveals transcription upstream of active human promoters. *Science*, **322**, 1851–1854.

Rao,S.S. *et al.* (2014) A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell*, **159**, 1665–1680.

Seila,A.C. *et al.* (2008) Divergent transcription from active promoters. *Science*, **322**, 1849–1851.

Sigova,A.A. *et al.* (2015) Transcription factor trapping by RNA in gene regulatory elements. *Science*, **350**, 978–981.

Sigova,A.A. *et al.* (2013) Divergent transcription of long noncoding RNA/mRNA gene pairs in embryonic stem cells. *Proc. Natl. Acad. Sci. USA*, **110**, 2876–2881.

Yang,L. *et al.* (2013) lncRNA-dependent mechanisms of androgen-receptor-regulated gene activation programs. *Nature*, **500**, 598–602.