Trends in Cell Biology

CellPress OPEN ACCESS

Review

(Alternative) transcription start sites as regulators of RNA processing

Carlos Alfonso-Gonzalez[1,2,3] and Valérie Hilgers[1,*]

Alternative transcription start site usage (ATSS) is a widespread regulatory strategy that enables genes to choose between multiple genomic loci for initiating transcription. This mechanism is tightly controlled during development and is often altered in disease states. In this review, we examine the growing evidence highlighting a role for transcription start sites (TSSs) in the regulation of mRNA isoform selection during and after transcription. We discuss how the choice of transcription initiation sites influences RNA processing and the importance of this crosstalk for cell identity and organism function. We also speculate on possible mechanisms underlying the integration of transcriptional and post-transcriptional processes.

Highlights

The alternative use of transcription start sites (TSSs) regulates the production of distinct mRNA isoforms, which have crucial roles in both physiological and pathological contexts.

Recent technological advances have provided deeper insights into the role of TSSs in the regulation of co-transcriptional and post-transcriptional RNA processing.

We discuss the impact and possible mechanisms of the crosstalk between transcription and co-transcriptional RNA processing, especially 3′-end site selection.

Introduction

In a given gene, the initiation of transcription can occur at different genomic locations, for example, in response to developmental or external signals. While some **TSSs** (see Glossary) lie a few nucleotides apart, thereby sharing *cis*-regulatory promoter sequences, others are distantly interspersed within the gene locus and exposed to distinct regulatory signals in *cis* and in *trans*. **ATSS** is widespread across metazoans (Figure 1A) and has a crucial role in regulating the tissue- or context-specific expression of genes. The formation of distinct transcript 5′-ends as a direct consequence of ATSS greatly impacts the diversity of transcript isoforms; in fact, the influence of ATSS on mRNA isoform diversity was shown to surpass that of **alternative splicing (AS)** in human tissues [1].

Consequences of ATSS on mRNA post-transcriptional regulation

One immediate consequence of ATSS is the selection of alternative **5′-untranslated regions (5′-UTRs)**, which may influence mRNA **post-transcriptional regulation**, for example, translation efficiency and stability (Figure 1B). The adoption of alternative 5′-UTRs, even if their sequence differs only minimally when alternative TSSs are closely spaced, can enhance translational efficiency by up to 100-fold [2]. **Polysome sequencing** experiments in mouse fibroblasts estimated that ~20% of multiple-TSS genes regulate translation through ATSS [3].

Alternative 5′-UTR-mediated translational regulation can occur through increased 5′-UTR length, inclusion of repressive elements, or inclusion of translationally repressive **upstream open reading frames (uORFs)**. ATSS-mediated uORF usage was shown to regulate the translation of mRNA in response to cellular stimuli. In *Arabidopsis thaliana*, exposing etiolated seedlings to blue light triggered widespread ATSS, inducing usage of, often nearby, TSSs, which overcame uORF-mediated inhibition, allowing mRNA translation in response to blue light [4]. During various stages of tumorigenesis, switching between 5′-UTR isoforms, each containing distinct post-transcriptional regulatory elements, including uORFs, **RNA-binding protein (RBP)** binding **motifs**, **5′-terminal oligopyrimidine (5′-TOP)** motifs, and **pyrimidine-rich translational elements (PRTEs)**, determines the translational potential of an mRNA. In a mouse model of squamous cell carcinomas, translated transcripts increasingly adopt 5′-TOP and PRTE motifs, typically have shorter

[1]Max Planck Institute of Immunobiology and Epigenetics, 79108 Freiburg, Germany
[2]Faculty of Biology, Albert Ludwigs University, 79104 Freiburg, Germany
[3]International Max Planck Research School for Molecular and Cellular Biology (IMPRS- MCB), 79108 Freiburg, Germany

*Correspondence: hilgers@ie-freiburg.mpg.de (V. Hilgers).

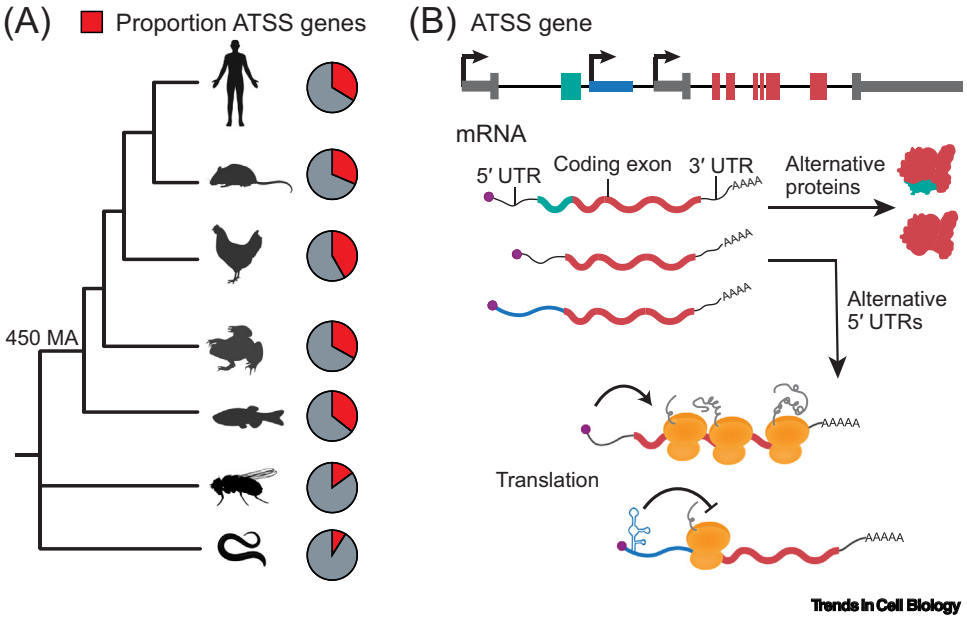


Figure 1. Widespread regulation through alternative transcription start site usage (ATSS). (A) Phylogenetic tree of various species, with branch lengths representing evolutionary animal divergence, based on Ensembl release 110 (July 2023). Pie charts depict the proportion of genes exhibiting ATSS in each group. Transcription start sites (TSSs) were considered distinct if located more than 50 nucleotides apart; therefore, ATSS here is concomitant with alternative use of *cis*-regulatory promoter elements. Species representing animal groups from top to bottom: *Homo sapiens*, *Mus musculus*, *Gallus gallus*, *Xenopus laevis*, *Danio rerio*, *Drosophila melanogaster*, and *Caenorhabditis elegans*. (B) ATSS-regulated gene with three TSSs (each denoted by a black arrow) from distinct promoters and differently colored alternative exons. Coding exons and untranslated regions (UTRs) are denoted by thick and thin boxes, respectively. ATSS can determine alternative first exons, influencing the generation of protein isoforms and/or RNA translation via regulatory elements in the 5′-UTRs. Abbreviation: MA, million years ago.

5′-UTRs, and exhibit reduced RNA secondary structures, all contributing to enhanced protein synthesis rates [5].

While the impact of 5′-UTR sequence variations on mRNA translation is well established, 5′-UTR-independent roles of ATSS, usually arising from differences in *cis*-regulatory regions, are not fully understood. Experiments in yeast, using reporter constructs featuring the same mRNA sequences but varying promoters, demonstrated that certain promoter sequences can enhance mRNA synthesis and degradation [6]. Single molecule imaging studies of cell cycle-regulated mRNA stability revealed that the promoter sequence alone can control **mRNA decay**, independent of the 5′ UTR, **3′-untranslated region (3′-UTR)**, or **open reading frame (ORF)** regions of the mRNA [7]. Intriguingly, promoters are essential for co-transcriptionally loading RBPs such as **Dbf2p** and **Rap1p**, which trigger mRNA decay in response to specific cellular cues [6,7]. This suggests a mechanism whereby transcription initiation and post-transcriptional regulation coordinate, affecting mRNA synthesis and decay rates, with ATSS having a key role in priming RNA processing in various cellular contexts.

ATSS in development and disease

During development, cells undergo substantial changes in their chromatin and transcriptional landscape, influenced by the availability of chromatin and **transcription factors (TFs)**. The maintenance of gene expression through critical developmental transitions relies on the utilization of multiple TSSs. In *Saccharomyces cerevisiae*, cell fate transitions are marked by differential expression from distinct TSSs, usually within distinct *cis*-regulatory environments, resulting in both

Glossary

3′-Untranslated region (3′-UTR): untranslated section of mRNAs located downstream of the coding sequence; involved in regulation of translation, stability, and localization.
5′–3′ Isoforms: mRNA isoforms distinguished by alternative boundaries at the 5′-UTR and/or 3′-UTR.
5′-Terminal oligopyrimidine (5′-TOP): sequence found at the 5′-end of some mRNAs; involved in the regulation of translation.
5′-Untranslated region (5′-UTR): untranslated section of mRNAs located upstream of the coding sequence; involved in regulation of translation and mRNA stability.
AAV-Perturb-seq: method designed for high-throughput screening and analysis of gene functions *in vivo* using adeno-associated virus (AAV) with the Perturb-seq methodology.
Alternative polyadenylation (APA): co-transcriptional mechanism that generates different mRNA isoforms through the selection of distinct polyadenylation sites in a gene, typically resulting in varying 3′-UTR lengths.
Alternative splicing (AS): co-transcriptional mechanism that generates different mRNA isoforms through differential selection of splice sites.
Alternative transcription start site usage (ATSS): usage of different TSSs within a gene; contributes to the diversity of mRNA transcripts.
C2H2: type of zinc finger protein domain characterized by two cysteines and two histidines binding a zinc ion.
CCCTC-binding factor (CTCF): highly conserved zinc finger protein that functions as a versatile transcriptional regulator; has a crucial role in gene expression, chromatin organization, and chromatin loop formation.
ChIP-seq: method for analyzing protein interactions with DNA by sequencing DNA fragments bound to specific proteins.
Cleavage and polyadenylation (CPA): process of cleaving the 3′-end of an RNA transcript and adding a poly (A) tail, crucial for mRNA stability and translation.
Cohesin: protein complex that regulates the separation of sister chromatids during cell division; also involved in gene regulation and DNA repair.
CREB-binding protein (CBP): transcriptional coactivator involved in the regulation of gene expression through

the co-expression and repression of TSSs from previous transitions [8]. In *Drosophila melanogaster*, developmentally regulated TFs and genes involved in tissue morphogenesis undergo significant ATSS throughout the life cycle, with >40% of developmental genes rapidly switching between at least two promoters [9].

ATSS during developmental transitions is thought to allow genes to adapt to specific chromatin and transcriptional landscapes. One of the most extensively studied cases illustrating this phenomenon is the **maternal-to-zygotic transition (MZT)**, considered one of the most significant changes in the transcriptome repertoire of the vertebrate life cycle. During MZT, a shift from maternal to zygotic transcription programs results in widespread ATSS and alterations in promoter architecture. Maternal-specific TSS selection, which necessitates an A/T-rich (W-box) motif, is replaced by the activation of zygotic TSSs characterized by broader patterns of dinucleotide enrichments. ATSS enables constitutively expressed genes to maintain their expression levels across the two distinct regulatory environments, illustrating that global ATSS events are a hallmark of development [10–12].

Functional studies have shed light on the significance of ATSS in the nervous system. For instance, in ***SCN1A***, which encodes the alpha subunit of the **NaV1.1** sodium channel and is associated with epilepsy and neurodevelopmental disorders, two TSSs from distinct promoters are co-expressed at similar levels, sharing the same ORF. The depletion of the **minor TSS** in *SCN1A* results in severe seizure phenotypes and a reduction in *SCN1A*expression. Interestingly, while the phenotype resulting from the depletion of the minor TSS is less severe than that associated with a full loss-of-function allele, the minor TSS mutants display seizure phenotypes under environmental stressors [13]. These findings underscore the critical function of ATSS in neuron development and functionality.

The widespread regulatory role of ATSS in various tissues and diseases becomes evident through transcriptome profiling studies, wherein TSS expression effectively discriminates among tissues, cell types, and cellular responses [14,15]. ATSS has been reported to be prevalent in cancer types, affecting a large fraction of cancer-associated genes; moreover, TSS choices can forecast the survival of patients with cancer, potentially offering a more precise prognostic tool compared with overall gene expression profiles [15,16].

Sites of coordination: transcription initiation guides co-transcriptional RNA processing

Contrary to the prevailing model that RNA processing relies exclusively on *cis*-regulatory elements within the RNA molecule and the *trans*-factors recruited to them, a growing body of evidence demonstrates the coordination of transcription initiation with co-transcriptional processes, resulting in discernible effects on isoform selection. This influence occurs in the context of AS and **alternative polyadenylation (APA)**, although we focus here on APA, the alternative use of 3′-end sites.

Long-read sequencing (LRS) techniques, capable of capturing full-length mRNA molecules, have significantly advanced our ability to identify and quantify coordinated co-transcriptional RNA processing events. Long reads allow for the detection and analysis of co-occurrence frequencies between exons within full-length mRNA molecules, and have revealed a widespread coordination between alternative TSSs, from distinct promoter, and alternative exons (**TSS-exon** coupling), as well as between TSSs and APA sites (**TSS-polyA** coupling) across various cell types, tissues, and developmental stages [17–19].

TSS-polyA coupling was recently shown to drive alternative 3′-end selection and the diversification of tissue-specific isoforms. Interestingly, the influence of TSS usage within the gene can be so

histone acetylation (also named p300/CBP).
Dbf2p: DNA-binding protein in yeast that functions as a serine/threonine-protein kinase; involved in the mitotic exit network and cell cycle regulation.
DNA-fluorescence *in situ* hybridization (FISH): technique used to detect and localize the presence or absence of specific DNA sequences on chromosomes.
Embryonic lethal abnormal vision (ELAV): member of a highly conserved family of RBPs involved in the regulation of mRNA processing and stability.
Enhancer RNAs (eRNAs): noncoding RNA molecules transcribed from enhancer regions; implicated in the regulation of gene expression.
Enhancer–promoter RNA interaction (EPRI): interactions between RNA molecules derived from enhancers and promoters; influence gene expression.
Exon-mediated activation of transcription starts (EMATS): process in which the active splicing of exons can activate weak upstream transcription start sites.
H2A.Z: histone variant involved in transcriptional regulation and chromatin dynamics.
H3K4me1: histone modification mark (monomethylation of lysine 4 on histone H3) associated with enhancer regions.
H3K18ac: histone modification (acetylation of lysine 18 on histone H3) associated with gene activation.
H3K36me3: histone modification mark (trimethylation of lysine 36 on histone H3) associated with actively transcribed genes.
Histone modifications: chemical modifications to histone proteins that affect chromatin structure and gene expression.
Long-read sequencing (LRS): sequencing technology that generates long reads; useful for the analysis of complex genomic regions and full-length transcript sequencing.
Maternal-to-zygotic transition (MZT): embryonic shift from dependence on maternal mRNA and protein products to the initiation of zygotic gene expression.
Methyl CpG binding protein 2 (MecP2): protein involved in DNA methylation and chromatin remodeling, with implications in neurological disorders.
Micro Capture-C (Micro-C): chromosome conformation capture method at the single nucleosome level; used to study chromatin architecture.

dominant that it overrides even the strongest polyadenylation signal, leading to the emergence of specific gene isoforms [19]. This TSS-polyA coordination process was termed 'promoter dominance' (Box 1). Recent studies described an additional mode of TSS-polyA coupling, known as the 'positional initiation-termination axis' (PITA). PITA suggests that the formation of TSS-polyA isoforms is constrained based on the order in which TSSs and 3′-end sites appear along the gene [20]. Single nuclei full-length isoform RNA sequencing in the human frontal cortex demonstrated that TSS-exon and TSS-polyA coordination displays greater cell-type specificity compared with exon–exon coordination. Interestingly, human-specific exons involving TSS-exon/polyA are coordinated as closely as highly conserved exons, suggesting that such coordination can quickly develop during the evolution of organism-specific celltypes [18].

Box 1. Alternative TSSs influence 3′-end site selection: promoter dominance

Promoter dominance is a gene regulatory mechanism whereby genes with ATSS selectively govern the expression of 3′-end isoforms in a TSS-specific manner. Such TSSs, termed 'dominant promoters', can over-ride *cis*-regulatory elements in the transcription unit, such as polyadenylation signals. Characterized by distinctive epigenetic marks and CBP binding, dominant promoters influence splice and polyadenylation choices and are commonly linked to a single transcript variant. *In vivo* alterations in dominant promoters through deletion or overexpression, and CBP mutations, induce significant shifts in 3′-end expression profiles. Notably, regulatory interactions between dominant promoters and associated 3′-end sites appear to be conserved, as indicated by coevolution between sequences in promoter regions and polyadenylation sites. This mutual information between distal regions hints at a *cis*-regulatory synergy between distal elements within a gene (Figure I).

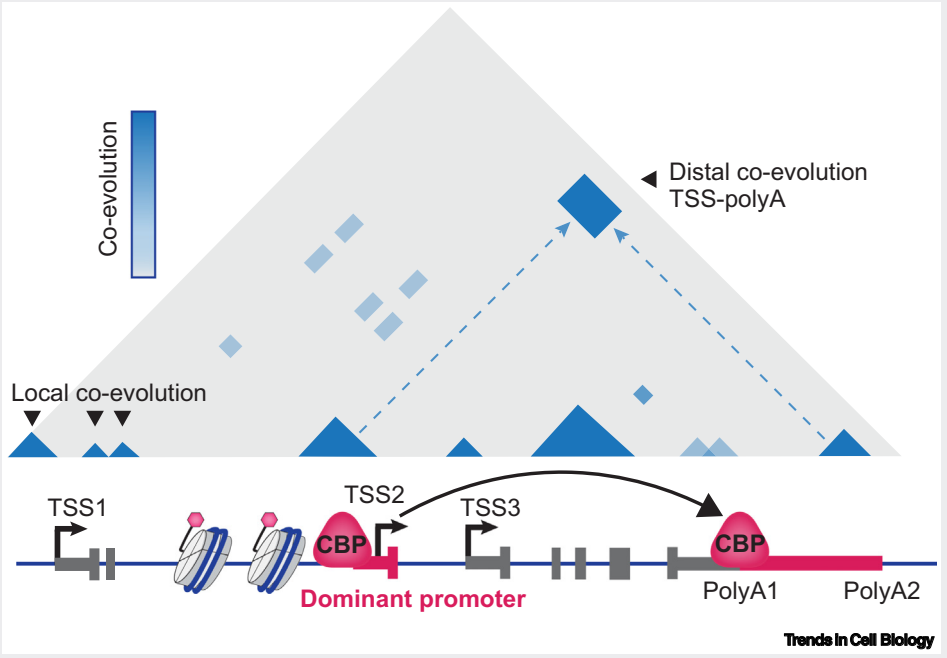
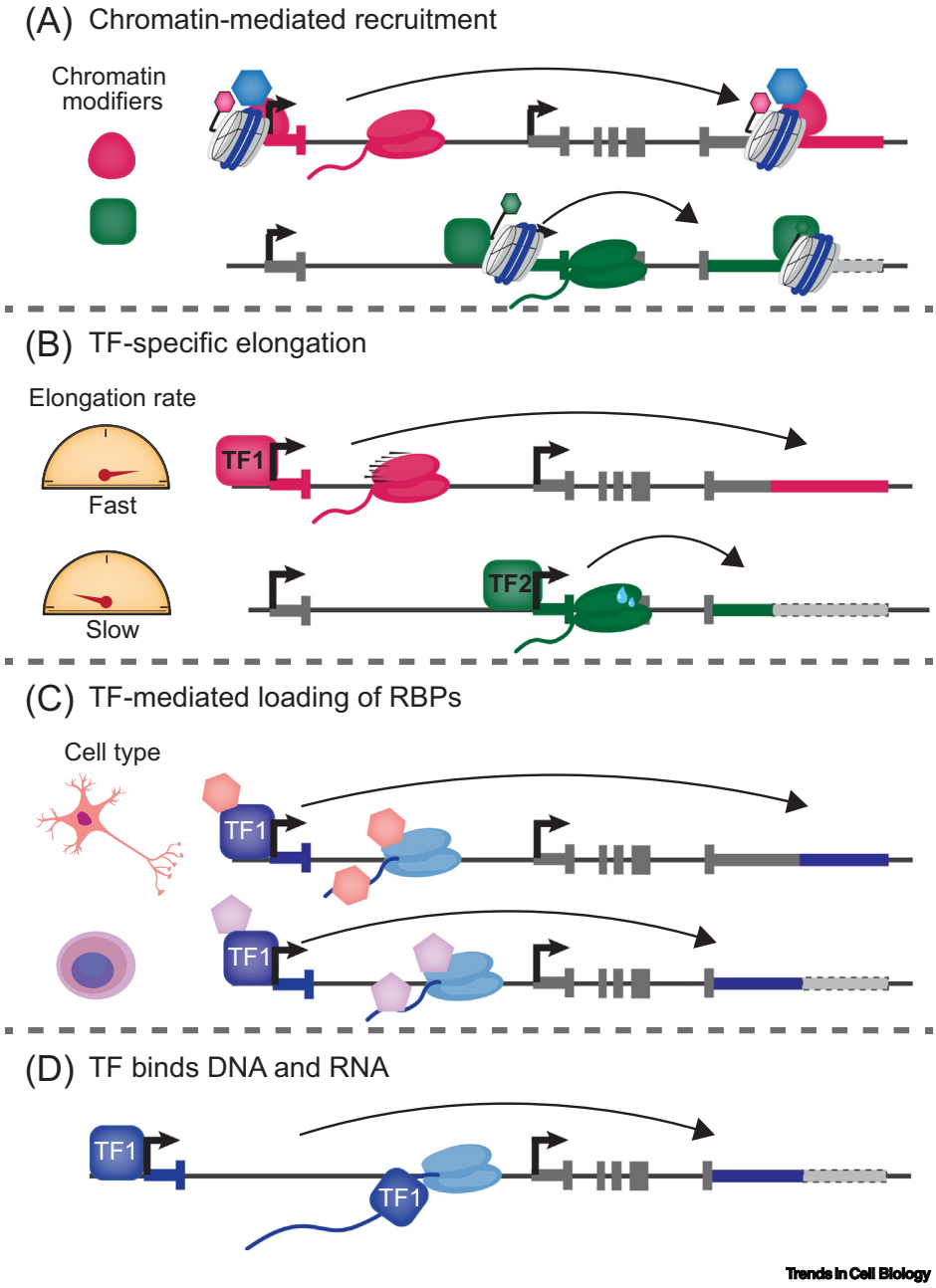


Figure I. Promoter dominance regulates 3′-end processing choices. Schematic depicting promoter dominance, in which specific transcription start sites (TSSs), through features such as CREB-binding protein (CBP) binding, exert influence over the selection of polyadenylation sites, contributing to transcript diversity. TSSs are denoted with black arrows. At the top, a coevolution matrix illustrates the inter-related evolution between the TSS and polyadenylation sites. Coevolution follows the notion that functional interactions between sequences are conserved; the coevolution score, based on a mutual information analysis, indicates whether nucleotide pairs mutate jointly to maintain genetic interactions. Arrowheads denote coevolving nucleotides in close proximity to each other (local coevolution) or located in distant genomic regions (distal coevolution). The intensity of the shading represents the degree of coevolution between sequence elements, with key areas highlighted by blue squares.

Minor TSS: TSS within a gene that fosters lower expression levels compared with other TSSs.
Motifs: short, recurrent patterns in DNA or RNA predicted to mediate a biological function, often serving as binding sites for proteins.
mRNA decay: cellular process of degrading mRNA molecules, regulating gene expression by determining mRNA lifespan and availability for translation.
NaV1.1: transmembrane protein forming a voltage-gated sodium channel;
implicated in neurological disorders.
Open reading frame (ORF): sequence of DNA or RNA that can be translated into a protein.
Perturb-seq: a genetic approach that combines CRISPR-based gene editing and screening with the analysis of phenotypes by single-cell RNA sequencing.
Polysome sequencing: technique to analyze the association of mRNAs with multiple ribosomes, providing insights into translation efficiency and regulation.
Post-transcriptional regulation: control of gene expression at the mRNA level, including processes such as transport, stability, and translation.
Pyrimidine-rich translational element (PRTE): RNA element that enhances mRNA translation.
Rap1p: DNA-binding protein involved in telomere length regulation and transcriptional activation.
RNA polymerase II (Pol II): enzymatic protein machinery responsible for transcribing mRNA from DNA in eukaryotic cells.
RNA-binding protein (RBP): protein that binds to RNA molecules; involved in various aspects of RNA metabolism.
***SCN1A*:** gene encoding the alpha subunit of the voltage-gated sodium channel NaV1.1, involved in neuronal function and disorders.
SET1 and SET2: histone methyltransferases involved in chromatin modification and gene regulation.
Sp1: zinc finger TF involved in the regulation of various genes.
Splicing: process of removing introns and joining exons in a pre-mRNA transcript to form a mature mRNA.
Topologically associating domains (TADs): genomic regions that interact more frequently with each other than with other regions; have a role in gene regulation.



Transcription factor (TF): protein that binds to specific DNA sequences to regulate transcription.
Transcription start site (TSS): site where the transcription of a gene begins.
TSS-exon: coordination between TSSs and exons in RNA processing.
TSS-polyA: coordination between TSSs and polyadenylation sites in RNA processing.
Upstream open reading frame (uORF): short ORF located upstream of the main ORF in mRNA, potentially regulating translation.

Figure 2. Mechanisms of alternative transcription start site usage-mediated coordination of RNA processing. (A) Chromatin-mediated recruitment: chromatin modifiers facilitate the recruitment of specific transcription factors (TFs) or RNA-binding proteins (RBPs; blue hexagons) to specific transcription start site (TSS) regions. A chromatin modifier targets a chromatin site, thereby promoting recruitment to this region. TSS-specific histone marks (green and pink hexagons) delineate the boundaries of the 5′–3′ isoform, thereby having a role in TSS-polyA coupling. (B) TF-specific RNA polymerase II (Pol II) elongation drives TSS-polyA couplings. Two scenarios are represented. Top: a putative TF (red) enhances the Pol II elongation rate, promoting read-through of the proximal poly(A) signal and transcription of the distal 3′-untranslated region (UTR;red). Bottom: a different TF (green) inhibits elongation speed and promotes cleavage and polyadenylation after the proximal 3′-UTR (green). (C) TF-mediated loading of RBPs: TFs recruit different RBPs in a celltype-dependent manner, driving tissue-specific TSS-polyA couplings. (D) TF binds DNA and RNA: a TF binds to DNA at the TSS and is then loaded onto the newly synthesized RNA, promoting the selection of the proximal 3′-untranslated region (UTR) (blue) in a TSS-specific manner.

Possible mechanisms mediating the coordination between transcription initiation and RNA processing

The widespread ATSS-mediated coordination of RNA processing points to a mechanism in which the promoter holds the information that primes the regulation of downstream events. Early studies revealed that altering specific promoter sequences increases **splicing** efficiency of fibronectin exon ED I by up to tenfold, with transcription levels remaining unchanged [21]. Similarly, specific promoters and enhancers can increase 3′-end processing activity of polyadenylation sites [22], and activation of tissue-specific promoters drive the expression of tissue-specific 3′-ends [19]. It is conceivable that the presence of multiple TSSs, each with its own set of *cis*-regulatory elements, provides the gene with the ability to accommodate multiple transcriptional states characterized by distinct chromatin structures, transcription elongation rates, and the recruitment of specific co-factors (Figure 2). Interestingly, genes with TSS-polyA tend to display a structural pairing of coupled start and end sites, as well as rapid **RNA polymerase II (Pol II)** elongation dynamics [20], hinting at multiple, possibly concomitant, mechanisms of TSS-polyA coupling.

Chromatin states shape ATSS-mediated co-transcriptional RNA processing

ATSS-mediated coordination of RNA processing poses a scenario in which multiple protein complexes and components interact throughout the transcription process, raising the question of how this continuous protein flow is accommodated at the chromatin level. Chromatin states influence RNA processing through a variety of mechanisms [23,24]. One significant manifestation of chromatin structure is its effect on Pol II dynamics. For instance, within the gene body, exons exhibit a denser nucleosome presence compared with introns, and this enrichment is directly linked to exon usage [25,26]. Consequently, Pol II moves faster through intronic regions compared with exonic ones, the variance in transcription speed providing key intervals for the recruitment of RNA-processing factors. This modulatory function of chromatin on TSS-exon couplings extends to the regulation of 3′-end processing. Polyadenylation sites are predominantly devoid of nucleosomes, but the areas downstream from these sites are heavily nucleosome enriched [23,27].

Beyond nucleosomal density, **histone modifications** regulate co-transcriptional RNA processing [28] and predict splicing outcomes in a positional and combinatorial manner [29,30]. The levels of **H3K4me1** and **H3K36me3** histone modifications show a strong association with polyadenylation site locations, indicating their possible influence on alternative 3′-end site selection [27]. Depletion of the highly conserved methyltransferases **SET1 and SET2** results in a widespread change in APA [31]. Genes with TSS-polyA coordination show a marked enrichment for **H3K18ac** and a notable depletion of the histone variant **H2A.Z**. Furthermore, the histone acetyltransferase **CREB-binding protein (CBP)** was found to be enriched at both the TSS and 3′-end of TSS-polyA coordinated genes [19]. Such characteristics have previously been linked to high transcriptional rates and robust **topologically associating domain (TAD)** insulation [32] (Figure 2A). Together, this recent evidence suggests that specific chromatin landscapes contribute to establishing a conducive environment for co-transcriptional RNA processing. Chromatin modifiers, such as CBP, may influence the nucleosomal landscape at the 5′ and 3′-end sites of specific gene isoforms, thereby enabling the targeted recruitment of regulatory factors in a chromatin-dependent manner (Figure 2A and see Figure I in Box 1). Further work is required to solve the mechanisms through which CBP, and likely other factors, link TSSs and 3′-end sites; for example, it remains to be assessed whether the deposition of histone marks functionally contributes to TSS-polyA, or is merely a consequence of CBP binding.

Transcription elongation tunes co-transcriptional RNA processing

Transcription elongation dynamics exert a profound influence on RNA-processing events, a phenomenon intertwined with cellular and tissue-specific complexities. Pol II exhibits variations in its elongation and pausing rates during transcription, dictating the pace of RNA synthesis. Remarkably, the rate of transcriptional elongation varies by several orders of magnitude among genes. However, these variations remain consistent across different cell types, suggesting that elongation is primarily determined by sequence elements and gene structure [33]. Pol II elongation dynamics significantly influence transcription termination and the selection of 3′-end isoforms. In mammalian cells, a Pol II variant characterized by a slow elongation rate leads to reduced Pol II accumulation within 3′-UTRs, while fast Pol II mutants exhibit an increased presence in distant genomic regions [34,35]. Similarly, in yeast, rapid Pol II mutants result in a notable shift in the generation of distal 3′-UTR isoforms [36]. In *Drosophila*, the introduction of a slow Pol II mutant affected polyadenylation site selection in 3–5% of genes, with a similar number displaying increased usage of polyadenylation sites either upstream or downstream [37].

The contribution of Pol II elongation dynamics to RNA processing bears substantial physiological consequences. Elongation speed increases with age, and interventions aimed at reducing Pol II speed result in enhanced lifespan in animals [38]. Increased elongation speed during aging also leads to splicing and transcript fidelity deficiencies, along with increased circular (circ)RNA production. In the nervous system, slow Pol II mutations exert their most pronounced effects on longer genes, with more substantial impacts on splicing and termination, an effect hypothesized to be related to the specific chromatin landscape of neuronal tissues [39]. Similar findings have emerged in the context of cancer progression, where elongation defects in long genes lead to aberrant 3′-end processing of DNA damage response transcripts, contributing to cancer progression [40,41]. Although the influence of ATSS and alternative promoter usage on Pol II speed has not been systematically studied, loading of distinct TFs at different TSSs is likely to have a significant role in Pol II speed.

TFs regulate co-transcriptional RNA processing

The ability of TSSs to specifically drive the expression of particular isoforms may be attributed to *cis*-regulatory elements within promoters that recruit TFs. Various transcription and elongation factors have been shown to regulate APA, independently of their transcriptional activity, with specific TFs displaying varied efficiencies in the regulation of 3′-end site selection [22]. One possibility is that ATSS provides an expanded platform for various TFs to bind. This can regulate isoform expression and affect transcriptional dynamics, such as transcription elongation, in a manner specific to each TSS (Figure 2B). Another model posits that TFs assist in the loading of tissue-specific RBPs. An example of this is the homeobox TF Ultrabithorax (Ubx) in *Drosophila*. Although Ubx is expressed across various cell lineages, the proteins it interacts with differ between these lineages, from chromatin modifiers to RNA-processing factors [42]. These observations are consistent with findings that a *Ubx* loss-of-function leads to RNA-processing defects, suggesting that effective RNA processing necessitates a protein that co-transcriptionally integrates *trans*-regulatory complexes [43] (Figure 2C).

Apart from their functions in transcription activation, some TFs modulate co-transcriptional RNA processing by binding to RNA [43–46]. Such regulatory mechanisms (Figure 2D) might be more prevalent than previously recognized, especially in light of recent discoveries that a substantial portion of TFs can bind to RNA [47]. An intriguing example is the **C2H2** zinc-finger protein **Sp1**, a ubiquitously expressed TF that regulates housekeeping genes. Sp1 binds to the long 3′-UTR of mRNAs and co-transcriptionally inhibits processing at distal polyadenylation sites by interacting with subunits of the core **cleavage and polyadenylation (CPA)** machinery, resulting in the degradation of distal 3′-UTR transcripts [44].

Gene looping and its role in defining boundaries for co-transcriptional RNA processing

TSS-polyA regulation is intriguing as it links two distant regions within the same gene. In yeast, genetic studies led to the proposal of a mechanism in which gene looping occurs between promoter regions and transcription end sites, thus establishing 5′–3′ boundaries [48–50]. Genetic studies have further demonstrated that gene looping necessitates the involvement of both transcription initiation and 3′-processing machineries [51]. Looping factors have also been implicated more generally in the regulation of transcription and RNA processing. For instance, the recruitment of **cohesin** by **CCCTC-binding factor (CTCF)** is essential for proper 3′-end processing, and the loss of CTCF results in the misregulation of APA [52] and AS [53]. Existing functional evidence indicates that factors and *cis*-regulatory elements involved in transcription initiation and termination are essential for gene looping [51]. However, alternate experimental approaches provide divergent evidence. **DNA-fluorescence *in situ* hybridization (FISH)** experiments on long, highly transcribed genes revealed that gene looping occurs during transcription, potentially allowing individual protein complexes to interact with Pol II co-transcriptionally as they travel along the gene axis. These loops, marked by chromatin condensation around the TSS and 3′-ends, are open rather than closed, challenging the traditional model of gene looping [54].

Alternative models propose that gene globules, rather than loops, mediate intragenic regulatory crosstalk. Recent advances in chromosome conformation capture technologies, particularly single nucleotide-resolution methods such as **Micro Capture-C (Micro-C)**, have shifted perspectives on gene structure by detecting gene globule structures, marked by extensive intragenic interactions and not by classic gene loops [55,56]. These globular structures may significantly affect interactions between intragenic regions and the TSS, influencing transcription regulation. The TF **methyl CpG binding protein 2 (MecP2)** exemplifies this in neuronal gene differentiation, where its recruitment to promoters, potentially facilitated by gene looping, modulates elongation dynamics and represses gene expression [57]. Recent work showed that genes with TSS-polyA coordination differ from other genes by forming gene looping domains [20]. The prediction of gene boundaries from chromosome conformation capture data revealed the existence of various compartments within the gene, which reflect the distinct TSS-polyA couplings. This suggests that such genes organize themselves into sections to support the expression of multiple TSS-polyA isoforms. These findings highlight the complexity of interpreting gene structure. Despite the differences, a common theme emerges: intragenic interactions with promoters serve as a feedback mechanism, relaying co-transcriptional events back to the promoter to regulate transcription.

Hitchhiking of RBPs at the onset of transcription

ATSS-mediated coordination of RNA processing could involve the recruitment of RBPs to TSSs. Large-scale **ChIP-seq** assays on RBPs have shown that RBPs associate with chromatin [58], with a notable preference for euchromatin and gene promoters. The recruitment of RBPs to the promoter region may be essential for proper RNA processing. In the developing nervous systems of flies and mammals, the RBP **embryonic lethal abnormal vision (ELAV)** regulates the expression of neuron-specific 3′-UTR isoforms. This was shown to depend on specific promoter elements [59], suggesting a mechanism in which RNA regulatory proteins are loaded from the outset, priming 3′-regulation right from transcription initiation [60]. It is possible that the recruitment of RBPs to specific promoters may be regulated through interactions with enhancers, for example, through binding **enhancer RNAs (eRNAs)**. Genome-wide estimates suggest that ~25% of enhancers produce eRNAs, which, according to **enhancer–promoter RNA interaction (EPRI)** studies, can influence looping selectivity by forming an RNA duplex through reverse complementarity with promoter RNAs [61]. Computational analyses indicated that RBPs can localize in overlapping regions near sites of TF binding [58], wherein enhancers could serve as platforms

for TF-RBP complexes to regulate isoform selection transcriptionally. Moreover, enhancers can regulate 3′-end processing activity and polyadenylation site selection, with distinct TFs displaying diverse efficiencies [22]. Further investigations are needed to test the functional involvement of RBPs in enhancer–promoter looping, and the RBP-chromatin and RBP-RNA interactions that underlie transcription regulatory processes.

Concluding remarks

Elucidating the interconnections between transcription initiation and RNA processing is critical for a deeper understanding of gene expression regulation. For example, AS has been shown to influence ATSS between distant TSSs through a process known as **exon-mediated activation of transcription starts (EMATS)**. In this mechanism, the active splicing of internal exons leads to the activation of cryptic, weak TSSs located 1 or 2kb upstream [62,63]. This finding contributed to the hypothesis that RNA processing serves as a regulatory checkpoint within the gene, modulating and controlling TSS activity. Such a process could foster a regulatory-loop mechanism internal to the gene for transcriptional regulation. Mechanisms such as EMATs underscore the significance of exploring co-transcriptional RNA processing to understand transcriptional control mechanisms.

Several aspects of the interaction between transcription initiation and RNA processing remain to be understood (see Outstanding questions). Recent research highlighted the critical role of *cis*-regulatory elements in transcriptional regulation, underscoring the functional importance of each nucleotide within these elements [64–66]. This raises important questions such ashow these regulatory sequences impact RNA processingand the extent of their variability in different tissue types and under various cellular conditions. An area for future exploration is also the potential involvement of RBPs or TFs with eRNAs. Their interaction could provide crucial insights into the mechanisms linking transcription and RNA dynamics.

Whether TSSs and 3′-end sites physically interact, and the manner in which these interactions influence the formation of transcriptional RNA processing complexes, is yet to be understood. The advent of high-resolution chromosome capture methods, such as Micro-C [67] and singlemolecule live imaging [68] could shed light on the dynamic interactions between intragenic elements. Dissecting the functional consequences of TSS-polyA associations is notably challenging, as it requires differentiating the specific roles of 5′-UTR and 3′-UTR regulatory regions. This complexity is compounded by previous findings that demonstrate the masking effect that each of these regulatory elements can have on RNA, potentially obscuring their individual contributions to gene regulation [69]. Recent breakthroughs in singlecell screening technologies, such as **Perturb-seq** [70] and the*in vivo* variant **AAV-Perturb-seq** [71], open opportunities for high-throughput screenings. These methods could facilitate targeted depletions of entire **5′–3′ isoforms**, or specific targeting of either the 5′-UTR or 3′-UTR, across various cellular contexts. This approach would enable a comprehensive assessment of molecular phenotypes, considering both cell- and gene structure-specific effects.

Outstanding questions

What is the impact of ATSS-mediated RNA processing on development and disease?

To what extent do TFs regulate RNA processing?

Does gene looping help coordinate the crosstalk between TSSs and polyadenylation sites?

What is the regulatory grammar at promoters that drives co-transcriptional RNA-processing choices?

Acknowledgments

We thank Hasan Can Ozbulut, Sarah Holec, Anton Hess, Sakshi Gorey, and Mengjin Shi for their insights. Our work was funded by the Max Planck Society, the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) project 500335138, and under Germany's Excellence Strategy (CIBSS—EXC-2189—project ID 390939984), and the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (grant agreement no. ERC-2018-STG-803258).

Declaration of interests

None declared by authors.

References

1. Reyes, A. and Huber, W. (2018) Alternative start and termination sites of transcription drive most transcript isoform differences across human tissues. *Nucleic Acids Res.* 46, 582–592
2. Rojas-Duran, M.F. and Gilbert, W.V. (2012) Alternative transcription start site selection leads to large differences in translation activity in yeast. *RNA* 18, 2299–2305
3. Wang, X. *et al.* (2016) Pervasive isoform-specific translational regulation via alternative transcription start sites in mammals. *Mol. Syst. Biol.* 12, 875
4. Kurihara, Y. *et al.* (2018) Transcripts from downstream alternative transcription start sites evade uORF-mediated inhibition of gene expression in Arabidopsis. *Proc. Natl. Acad. Sci. USA* 115, 7831–7836
5. Weber, R. *et al.* (2023) Monitoring the 5'UTR landscape reveals isoform switches to drive translational efficiencies in cancer. *Oncogene* 42, 638–650
6. Bregman, A. *et al.* (2011) Promoter elements regulate cytoplasmic mRNA decay. *Cell* 147, 1473–1483
7. Trcek, T. *et al.* (2011) Single-molecule mRNA decay measurements reveal promoter-regulated mRNA stability in yeast. *Cell* 147, 1484–1497
8. Chia, M. *et al.* (2021) High-resolution analysis of cell-state transitions in yeast suggests widespread transcriptional tuning by alternative starts. *Genome Biol.* 22, 34
9. Batut, P. *et al.* (2013) High-fidelity promoter profiling reveals widespread alternative promoter usage and transposon-driven developmental gene expression. *Genome Res.* 23, 169–180
10. Haberle, V. *et al.* (2014) Two independent transcription initiation codes overlap on vertebrate core promoters. *Nature* 507, 381–385
11. Zhang, P. *et al.* (2017) Relatively frequent switching of transcription start sites during cerebellar development. *BMC Genomics* 18, 461
12. Cvetesic, N. *et al.* (2020) Global regulatory transitions at core promoters demarcate the mammalian germline cycle. *bioRxiv*, Published online November 20, 2020. https://doi.org/10.1101/2020.10.30.361865
13. Haigh, J.L. *et al.* (2021) Deletion of a non-canonical regulatory sequence causes loss of Scn1a expression and epileptic phenotypes in mice. *Genome Med.* 13, 69
14. Cass, A.A. and Xiao, X. (2019) mountainClimber identifies alternative transcription start and polyadenylation sites in RNA-seq. *Cell Syst.* 9, 393–400
15. Demircioğlu, D. *et al.* (2019) A pan-cancer transcriptome analysis reveals pervasive regulation through alternative promoters. *Cell* 178, 1465–1477
16. Huang, K.K. *et al.* (2021) Long-read transcriptome sequencing reveals abundant promoter diversity in distinct molecular subtypes of gastric cancer. *Genome Biol.* 22, 44
17. Anvar, S.Y. *et al.* (2018) Full-length mRNA sequencing uncovers a widespread coupling between transcription initiation and mRNA processing. *Genome Biol.* 19, 46
18. Hardwick, S.A. *et al.* (2022) Single-nuclei isoform RNA sequencing unlocks barcoded exon connectivity in frozen brain tissue. *Nat. Biotechnol.* 40, 1082–1092
19. Alfonso-Gonzalez, C. *et al.* (2023) Sites of transcription initiation drive mRNA isoform selection. *Cell* 186, 2438–2455
20. Calvo-Roitberg, E. *et al.* (2024) mRNA initiation and termination are spatially coordinated. *bioRxiv*, Published online January 7, 2024. https://doi.org/10.1101/2024.01.05.574404
21. Cramer, P. *et al.* (1997) Functional association between promoter structure and transcript alternative splicing. *Proc. Natl. Acad. Sci. USA* 94, 11456–11460
22. Kwon, B. *et al.* (2022) Enhancers regulate 3' end processing activity to control expression of alternative 3'UTR isoforms. *Nat. Commun.* 13, 2709
23. Proudfoot, N.J. (2016) Transcriptional termination in mammals: stopping the RNA polymerase II juggernaut. *Science* 352, aad9926
24. Tellier, M. *et al.* (2020) Transcription and splicing: a two-way street. *Wiley Interdiscip. Rev. RNA* 11, e1593
25. Schwartz, S. *et al.* (2009) Chromatin organization marks exon-intron structure. *Nat. Struct. Mol. Biol.* 16, 990–995
26. Tilgner, H. *et al.* (2009) Nucleosome positioning as a determinant of exon recognition. *Nat. Struct. Mol. Biol.* 16, 996–1001
27. Spies, N. *et al.* (2009) Biased chromatin signatures around polyadenylation sites and exons. *Mol. Cell* 36, 245–254
28. Luco, R.F. *et al.* (2010) Regulation of alternative splicing by histone modifications. *Science* 327, 996–1000
29. Agirre, E. *et al.* (2021) Splicing-associated chromatin signatures: a combinatorial and position-dependent role for histone marks in splicing definition. *Nat. Commun.* 12, 682
30. Xu, S.-J. *et al.* (2021) Chromatin-mediated alternative splicing regulates cocaine-reward behavior. *Neuron* 109, 2943–2966
31. Kaczmarek Michaels, K. *et al.* (2020) Regulation of alternative polyadenylation in the yeast *Saccharomyces cerevisiae* by histone H3K4 and H3K36 methyltransferases. *Nucleic Acids Res.* 48, 5407–5425
32. Ibarra-Morales, D. *et al.* (2021) Histone variant H2A.Z regulates zygotic genome activation. *Nat. Commun.* 12, 7002
33. Veloso, A. *et al.* (2014) Rate of elongation by RNA polymerase II is associated with specific gene features and epigenetic modifications. *Genome Res.* 24, 896–905
34. Fong, N. *et al.* (2015) Effects of transcription elongation rate and Xrn2 exonuclease activity on RNA polymerase II termination suggest widespread kinetic competition. *Mol. Cell* 60, 256–267
35. Cortazar, M.A. *et al.* (2019) Control of RNA Pol II speed by PNUTS-PP1 and Spt5 dephosphorylation facilitates termination by a 'sitting duck torpedo' mechanism. *Mol. Cell* 76, 896–908
36. Geisberg, J.V. *et al.* (2020) The transcriptional elongation rate regulates alternative polyadenylation in yeast. *eLife* 9, e59810
37. Liu, X. *et al.* (2017) Transcription elongation rate has a tissue-specific impact on alternative cleavage and polyadenylation in *Drosophila melanogaster*. *RNA* 23, 1807–1816
38. Debès, C. *et al.* (2023) Ageing-associated changes in transcriptional elongation influence longevity. *Nature* 616, 814–821
39. Maslon, M.M. *et al.* (2019) A slow transcription rate causes embryonic lethality and perturbs kinetic coupling of neuronal genes. *EMBO J.* 38, e101244
40. Miller, T.E. *et al.* (2017) Transcription elongation factors represent in vivo cancer dependencies in glioblastoma. *Nature* 547, 355–359
41. Krajewska, M. *et al.* (2019) CDK12 loss in cancer cells affects DNA damage response genes through premature cleavage and polyadenylation. *Nat. Commun.* 10, 1757
42. Carnesecchi, J. *et al.* (2020) Multi-level and lineage-specific interactomes of the Hox transcription factor Ubx contribute to its functional specificity. *Nat. Commun.* 11, 1388
43. Carnesecchi, J. *et al.* (2022) The Hox transcription factor Ultrabithorax binds RNA and regulates co-transcriptional splicing through an interplay with RNA polymerase II. *Nucleic Acids Res.* 50, 763–783
44. Song, J. *et al.* (2022) Regulation of alternative polyadenylation by the C2H2-zinc-finger protein Sp1. *Mol. Cell* 82, 3135–3150
45. Zhu, L. *et al.* (2022) Transcription factor GATA4 regulates cell type-specific splicing through direct interaction with RNA in human induced pluripotent stem cell-derived cardiac progenitors. *Circulation* 146, 770–787
46. Jin, B. *et al.* (2023) MEN1 is a regulator of alternative splicing and prevents R-loop-induced genome instability through suppression of RNA polymerase II elongation. *Nucleic Acids Res.* 51, 7951–7971
47. Oksuz, O. *et al.* (2023) Transcription factors interact with RNA to regulate genes. *Mol. Cell* 83, 2449–2463
48. Ansari, A. and Hampsey, M. (2005) A role for the CPF 3'-end processing machinery in RNAP II-dependent gene looping. *Genes Dev.* 19, 2969–2978
49. O'Sullivan, J.M. *et al.* (2004) Gene loops juxtapose promoters and terminators in yeast. *Nat. Genet.* 36, 1014–1018
50. Tan-Wong, S.M. *et al.* (2012) Gene loops enhance transcriptional directionality. *Science* 338, 671–675
51. Hampsey, M. *et al.* (2011) Control of eukaryotic gene expression: Gene loops and transcriptional memory. *Adv. Enzym. Regul.* 51, 118–125
52. Nanavaty, V. *et al.* (2020) DNA methylation regulates alternative polyadenylation via CTCF and the cohesin complex. *Mol. Cell* 78, 752–764

53. Alharbi, A.B. *et al.* (2021) CTCF as a regulator of alternative splicing: new tricks for an old player. *Nucleic Acids Res.* 49, 7825–7838
54. Leidescher, S. *et al.* (2022) Spatial organization of transcribed eukaryotic genes. *Nat. Cell Biol.* 24, 327–339
55. Hsieh, T.H.S. *et al.* (2016) Micro-C XL: assaying chromosome conformation from the nucleosome to the entire genome. *Nat. Methods* 13, 1009–1011
56. Hsieh, T.H.S. *et al.* (2015) Mapping nucleosome resolution chromosome folding in yeast by Micro-C. *Cell* 162, 108–119
57. Boxer, L.D. *et al.* (2020) MeCP2 represses the rate of transcriptional initiation of highly methylated long genes. *Mol. Cell* 77, 294–309
58. Van Nostrand, E.L. *et al.* (2020) A large-scale binding and functional map of human RNA-binding proteins. *Nature* 583, 711–719
59. Oktaba, K. *et al.* (2015) ELAV links paused pol II to alternative polyadenylation in the drosophila nervous system. *Mol. Cell* 57, 341–348
60. Hilgers, V. (2015) Alternative polyadenylation coupled to transcription initiation: Insights from ELAV-mediated 3' UTR extension. *RNA Biol.* 12, 918–921
61. Liang, L. *et al.* (2023) Complementary Alu sequences mediate enhancer–promoter selectivity. *Nature* 619, 868–875
62. Fiszbein, A. *et al.* (2019) Exon-mediated activation of transcription starts. *Cell* 179, 1551–1565
63. Uriostegui-Arcos, M. *et al.* (2023) Splicing activates transcription from weak promoters upstream of alternative exons. *Nat. Commun.* 14, 3435
64. Fuqua, T. *et al.* (2020) Dense and pleiotropic regulatory information in a developmental enhancer. *Nature* 587, 235–239
65. Avsec, Ž. *et al.* (2021) Base-resolution models of transcription-factor binding reveal soft motif syntax. *Nat. Genet.* 53, 354–366
66. Bergman, D.T. *et al.* (2022) Compatibility rules of human enhancer and promoter sequences. *Nature* 607, 176–184
67. Hamley, J.C. *et al.* (2023) Determining chromatin architecture with Micro Capture-C. *Nat. Protoc.* 18, 1687–1711
68. Mach, P. *et al.* (2022) Cohesin and CTCF control the dynamics of chromosome folding. *Nat. Genet.* 54, 1907–1918
69. Theil, K. *et al.* (2018) Post-transcriptional regulation by 3' UTRs can be masked by regulatory elements in 5' UTRs. *Cell Rep.* 22, 3217–3226
70. Schraivogel, D. *et al.* (2020) Targeted Perturb-seq enables genome-scale genetic screens in single cells. *Nat. Methods* 17, 629–635
71. Santinha, A.J. *et al.* (2023) Transcriptional linkage analysis with in vivo AAV-Perturb-seq. *Nature* 622, 367–375