

RNA helicase-dependent gene looping impacts messenger RNA processing

Sophie Terrone^{1,†}, Jessica Valat^{1,†}, Nicolas Fontrodona¹, Guillaume Giraud^{1,†}, Jean-Baptiste Claude¹, Emmanuel Combe^{1,†}, Audrey Lapendry¹, H el ene Polv eche^{1,2}, Lamya Ben Ameer¹, Arnaud Duvermy¹, Laurent Modolo¹, Pascal Bernard¹, Franck Mortreux¹, Didier Auboeuf¹ and Cyril F. Bourgeois^{1,*}

¹Laboratoire de Biologie et Mod elisation de la Cellule, Ecole Normale Sup erieure de Lyon, CNRS, UMR 5239, Inserm, U1293, Universit e Claude Bernard Lyon 1, 46 all e d'Italie, F-69364 Lyon, France and ²CECS/AFM, I-STEM, 28 rue Henri Desbru eres, F-91100, Corbeil-Essonnes, France

Received September 10, 2021; Revised July 25, 2022; Editorial Decision August 03, 2022; Accepted August 25, 2022

ABSTRACT

DDX5 and DDX17 are DEAD-box RNA helicase paralogs which regulate several aspects of gene expression, especially transcription and splicing, through incompletely understood mechanisms. A transcriptome analysis of DDX5/DDX17-depleted human cells confirmed the large impact of these RNA helicases on splicing and revealed a widespread deregulation of 3' end processing. *In silico* analyses and experiments in cultured cells showed the binding and functional contribution of the genome organizing factor CTCF to chromatin sites at or near a subset of DDX5/DDX17-dependent exons that are characterized by a high GC content and a high density of RNA Polymerase II. We propose the existence of an RNA helicase-dependent relationship between CTCF and the dynamics of transcription across DNA and/or RNA structured regions, that contributes to the processing of internal and terminal exons. Moreover, local DDX5/DDX17-dependent chromatin loops spatially connect RNA helicase-regulated exons with their cognate promoter, and we provide the first direct evidence that *de novo* gene looping modifies alternative splicing and polyadenylation. Overall our findings uncover the impact of DDX5/DDX17-dependent chromatin folding on pre-messenger RNA processing.

INTRODUCTION

Eukaryotic transcription by RNA polymerase II (RNAPII) produces pre-messenger RNA (pre-mRNA) that undergo different processing steps, including 5' capping, splicing and 3' end cleavage and polyadenylation. These steps mostly take place as the nascent transcript is still associated with the RNAPII complex, although a fraction of introns can be removed post-transcriptionally (1). This crosstalk between transcription and RNA processing involves many factors and works in both ways (reviewed in (2–4)).

For example, alternative splicing is affected by variations in the RNAPII elongation speed (5–11) and conversely, splicing increases transcription efficiency and was associated, albeit not systematically, with RNAPII pausing (12–20). The modulation of elongation rate was initially proposed to give a 'window of opportunity' for the recognition of some exons (7,21), but more recent evidence has shown that a simple kinetic model cannot explain all the alternative splicing changes. In fact, both fast and slow RNAPII have a mixture of positive and negative effects on exon inclusion (8,22). Other factors such as the folding of the nascent RNA or the structure and organization of chromatin also modulate the crosstalk between transcription and splice site selection (23,24).

Similarly, 3' end processing of RNAPII transcripts is directly linked to transcription termination (25), and it has also been shown to stimulate transcription initiation (26). Extensive interactions have been described between promoter-associated factors and cleavage/polyadenylation or termination factors (reviewed in (2)). These physical contacts between factors that associate to both distal ends of genes can result in a modification of gene architecture to form so-called gene loops. Gene looping is a widely

*To whom correspondence should be addressed. Tel: +33 47272 8663; Fax: +33 47272 8674; Email: cyril.bourgeois@inserm.fr

[†]The authors wish it to be known that, in their opinion, the first two authors should be regarded as Joint First Authors.

Present addresses:

Guillaume Giraud, Centre de Recherche en Canc erologie de Lyon (CRCL), Lyon, France.

Emmanuel Combe, Centre de Recherche en Canc erologie de Lyon (CRCL), Lyon, France.

described phenomenon in yeast, it is transcription-dependent and involves various factors (4,27–31). Only a handful of gene loops has been demonstrated so far in the human genome, but a common picture that has emerged from those studies is the highly dynamic nature of this phenomenon, as gene loops form or disappear in relation with gene transcription activation or repression (32–36).

Gene looping has been proposed to promote transcription memory and promoter directionality (37–39), as reviewed recently (40). It has also a role in delimitating the 3' end of transcription units, in ensuring termination, and in the regulation of alternative polyadenylation in yeast (28,34,41). In mammals, gene loops have been associated to the binding of the CCCTC-binding factor, CTCF (32,33), which plays important roles in 3D genome organization, together with the Cohesin complex (42–44). CTCF is mostly known to regulate transcription activation and repression via enhancer/promoter looping or insulation, but it has also been shown to regulate alternative splicing (45–49) and alternative polyadenylation (50).

In a recent study, CTCF-dependent chromatin looping was proposed to promote the usage of the polyadenylation site (PAS) that is located upstream of the CTCF binding site (50). Early evidence about the role of CTCF in alternative splicing also pointed out a similar position-dependent effect of CTCF relative to the targeted exon (45), and a correlation was established between the existence of CTCF/Cohesin-dependent chromatin loops joining alternative exons to their related promoter and the inclusion or skipping of those exons (51–54). However, whether chromatin looping impacts or not alternative splicing at the RNA level remains to be demonstrated.

DDX5 and its closely related paralog DDX17 belong to the large family of evolutionarily conserved DEAD-box ATP-dependent RNA helicases (55). These two proteins have largely redundant functions in various aspects of RNA metabolism, including pre-mRNA and microRNA processing (56). Previous work, from our lab and others, has shown that DDX5 and DDX17 control the inclusion of a large number of exons by modulating the folding of their target transcripts, thanks to their helicase activity, and by modulating the binding of splicing regulators to RNA (57–63). Recently, a new function of DDX5 and DDX17 in transcription termination has been reported, but the exact mechanism by which they do so is still uncertain (64,65). Both factors are thought to resolve R-loops downstream of the PAS, thereby promoting RNAPII release. However, the budding yeast DDX5 ortholog, Dbp2, also regulates transcription termination of a subset of protein-coding genes, but this activity rather seems to be correlated with the formation of secondary structures within the 3'UTR of targeted transcripts (66).

Importantly, beside their direct effects on RNA molecules, DDX17 and DDX5 also exert some of their functions at the chromatin level (67). DDX5 and DDX17 interact with CTCF and Cohesin, and DDX5 regulates CTCF insulator function (68,69). Both helicases also interact with many other transcription factors and epigenetic regulators (67,70), with consequences on gene expression that vary depending on the nature of the complex and on the cellular context. However, how

those chromatin-associated functions are achieved is unclear.

Here, we report that *DDX5* and *DDX17* silencing alters the inclusion of thousands of exons, and that a subset of those internal and 3' terminal exons is also regulated by CTCF. We show that CTCF binds to chromatin in a *DDX5/DDX17*-dependent manner downstream of those exons, which display a high RNAPII density and a high GC content. We propose the existence of an RNA helicase-dependent relationship between CTCF and the dynamics of transcription across DNA and/or RNA structured regions, that contributes to the processing of a subset of exons. We present evidence that local genes loops often spatially join *DDX5/DDX17*-regulated exons and their cognate promoter, and that *DDX5/DDX17* silencing alters gene looping. Finally, we demonstrate for the first time and on two different genes that creating a *de novo* promoter-exon loop has an impact on the inclusion of the looped exon at the RNA level.

MATERIALS AND METHODS

Cell culture and transfections

Human SH-SY5Y neuroblastoma cells (ECACC) were grown and transfected essentially as described previously (62). For standard *DDX5/DDX17* silencing experiments, 20 nM of siRNA was used and cells were harvested 48h later. For double knock-down experiments (*DDX5/DDX17* + CTCF), we used a total of 50 nM siRNA (20 nM si*DDX5/DDX17* + 30 nM siCTCF) and cells were harvested 36 h later, as CTCF depletion induced more cell lethality. SiCtrl: CGUACGCG-GAAUACUUCGA[dT][dT]; si*DDX5/DDX17*: GGCUA-GAUGUGGAAGAUGU[dT] [dT]; the siCTCF was an equimolar mixture of two different siRNAs (GAUAAGAC-CUCCGCCAGA[dT][dT] and AGAGGAAUCUUCU-UUCUUAGAGCGC[dT][dT]).

For treatment with transcription inhibitors (purchased from Merck-Millipore), SH-SY5Y cells were plated in 6-well-plates to reach 70% confluency, and then treated for 6 h with 300 nM Flavopiridol, 50 or 100 μ M 5,6-dichlorobenzimidazole 1- β -D-ribofuranoside (DRB) or Dimethylsulfoxide as a control, prior to RNA extraction.

To generate the SH3TC1- Δ CTCF stable cell line, we cloned the sequences corresponding to the two RNA guides into the *BsmBI* site of the CRIZI plasmid (provided by Philippe Mangeot, CIRI, Lyon). SH-SY5Y cells were then transfected with 1 μ g of xCas9 3.7 plasmid (Addgene #108379) and 1 μ g of gRNA-containing plasmid (500 ng of each guide) using jetPRIME (PolyPlus Transfection). The next day, cells were placed under selection pressure for 5 days in the presence of 400 μ g/ml G418 (Sigma-Aldrich G8168). After 2 weeks of cell growth, isolated colonies were selected, amplified, individually screened by PCR to check out the presence of the deletion, and finally sequenced to confirm the modification.

CLOuD9

The CLOuD9 design was adapted from the original study (71) with the following modifications. The HEK-CLOuD9

cell line was generated to stably express *Staphylococcus pyogenes* (SP) and *S. aureus* nuclease (SA) deficient Cas9, respectively fused to the dimerizable PYL1 and ABI1 domains, but devoid of their guide RNA (gRNA) scaffold sequence. The U6 promoters and gRNA scaffold sequences were deleted from both parental plasmids EF1a-dCas9-PYL1-P2A-Hygro and dCas9-ABI-P2A-Puro lentiviral cloning vectors (System Biosciences SBI, obtained from Dr Kevin C. Wang, Stanford University School of Medicine). The resulting lentiviral constructs were used to infect parental HEK293 cells (ATCC) at a MOI of 1.5, followed by puromycin and hygromycin selection for a week. Genomic integration was confirmed by PCR and protein expression of dCas9 was validated by Western-blotting using HA/Flag antibodies (data not shown). U6 promoters and gRNA scaffold sequences specific for SA and SP dCas9 were respectively subcloned from the original dCas9-ABI-P2A-Puro plasmid and from the PX459 plasmid (Addgene #62988) into pBluescript SK (–), between *KpnI* and *EcoRI* sites. This allowed to clone gRNA sequences into pBSK-SA (*BsmBI* site) and pBSK-SP (*BbsI* site). Sequences of gRNAs were the following: *FBLN1* promoter (AGAGACC-CGGGAAGTCACCG) and polyadenylation site (CAGT-GAAATGCTCACCTCCG), and *EYA3* promoter (CC-GAAACAGTGTGCACGAA) and exon 7 (TAAACCA-CACTCCTAAGCTG). Expression of gRNAs was confirmed by RT-PCR using primers located into gRNA sequence and gRNA scaffold (data not shown).

HEK-CLOuD9 cells were transfected with 1 µg of each gRNA-containing plasmid using jetPRIME (PolyPlus Transfection), 1 mM Abscisic Acid (Merck Sigma-Aldrich) or DMSO was added to the cells 4 h after transfection. Chromatin loop formation and RNA processing were monitored 48 h later, respectively by 3C and RT-qPCR.

Chromosome conformation capture (3C)

Cells (15×10^6 per assay) were fixed for 10 min with 1% formaldehyde (ThermoFisher). Cross-linking was stopped by addition of PBS-glycine for 5 min, cells were washed twice with PBS, harvested and incubated for 1 h on ice in lysis buffer (10 mM Tris pH 8.0, 100 mM NaCl, 2% NP-40 and protease inhibitor, Roche #11697498001). The cytoplasm was eliminated by centrifugation and nuclei were washed once in 1.25× digestion buffer (ThermoFisher #B64) and then resuspended in 250 µl of the same buffer. SDS was added to a final concentration of 0.8% and nuclei were mixed for 1 h at 37°C at 900 rpm in a Thermomixer® (Eppendorf) to permeabilize their membrane. Triton (final concentration 4%) was added together with another 250 µl of 1.25× digestion buffer to neutralize SDS, the mixture was incubated again for 1 h at 900 rpm at 37°C. Digestion was carried out by the successive addition of 3×20 µl of Fast-Digest *SacI* (for the *PRMT2* gene) or *PstI* (for the *NCSI* gene) enzyme (ThermoFisher), during a 24 h-long incubation at 37°C under mixing at 900 rpm. The overall efficiency of digestion was verified on a small aliquot by agarose gel electrophoresis. Digestion at specific restriction sites was also verified by qPCR using primers flanking the restriction site. This control showed that the digestion was homoge-

neous across the regions of interest and in all experimental conditions (data not shown).

Nuclei were then centrifuged at 2000 rpm for 10 min, washed once with 1× ligation buffer (ThermoFisher #B69) and resuspended in the same buffer. Ligation was performed overnight at 16°C under mixing at 750 rpm in the presence of 1 mM ATP and 250 U of T4 DNA Ligase (ThermoFisher), and continued for another 4 h following the addition of 175 U ligase and 1 mM ATP. The efficiency of ligation of a small aliquot was verified by agarose gel electrophoresis. After centrifugation at 3000 rpm for 10 min, nuclei were resuspended in water and cross-link was reversed by incubation with proteinase K (Roche) over night at 65°C under mixing at 1,500 rpm. DNA was purified first by phenol/chloroform extraction followed by isopropanol precipitation, and then on Agencourt AMPure XP beads (Beckman Coulter) to remove residual ATP.

Primers for 3C analyses (Supplementary Table S4) were designed to obtain amplicons of about 350 bp, and they were first tested and validated on bacterial artificial chromosomes (BACs) containing the regions of interest (RP11-1000I21 and RP11-788J10, ThermoFisher). BACs were digested and ligated as described for genomic DNA, and serial DNA dilutions were used for qPCR using the SYBR® Premix Ex Taq (Tli RNaseH Plus) (Takara), in the following conditions: 30 s of denaturation at 95°C, followed by 45× (95°C for 10 s, 65°C for 20 s). We selected only primer couples that gave a linear amplification during the PCR reaction, a Ct value within a similar range, and whose melting curve did not reveal any non-specific product. PCR products were also analysed by agarose gel electrophoresis to verify their size and purity. Primers were next validated on 80 ng of 3C samples to determine the exponential amplification range and to define the number of cycles to use for quantitative analyses (36 cycles for *NCSI*, 34 cycles for *PRMT2*). Amplified 3C products were analysed by 2% agarose gel electrophoresis and quantified in a GelDoc XR+ imager (Bio-Rad) using the Image Lab software. The amount of each 3C product was normalized to the loading control corresponding to a PCR product amplified between two restrictions sites (*GAPDH* for 3C-*PstI*, *DDX17* for 3C-*SacI*). Note that for 3C analyses of *FBLN1* and *EYA3* genes, that were made only to validate the formation of the loop induced by CLOuD9, the validation of the primers was limited to the standard determination of the amplification range and purity of the amplicon. Therefore, the signal observed at a given 3C fragment cannot be directly compared to the signal of another fragment, and we chose to represent the results of Figure 6 as a histogram, and not as a line connecting the different fragments.

Co-immunoprecipitation and western blotting

Total protein extraction was carried out as previously described (60). Primary antibodies used for western-blotting were: DDX5 (ab10261, Abcam), DDX17 (ab24601, Abcam), CTCF (ab128873, Abcam), Actin (sc-1616, SantaCruz), GAPDH (sc-32233, SantaCruz), CPSF5 (sc-81109, SantaCruz), CPSF6 (75169, Cell Signaling Technology).

For co-immunoprecipitation, SH-SY5Y cells were harvested and gently lysed for 5 min on ice in a buffer containing 10 mM Tris-HCl pH 8.0, 140 mM NaCl, 1.5 mM MgCl₂, 10 mM EDTA, 0.5% NP40, completed with protease and phosphatase inhibitors (Roche #11697498001 and #5892970001), to isolate the nuclei from the cytoplasm. After centrifugation, the nuclei were lysed in the IP buffer (20 mM Tris-HCl pH 7.5, 150 mM NaCl, 2 mM EDTA, 1% NP40, 10% glycerol and protease/phosphatase inhibitors) for 30 min at 4°C under constant mixing. The nuclear lysate was centrifuged for 15 min to remove debris and soluble proteins were quantified by BCA (ThermoFisher). The lysate was pre-cleared with 30 µl of Dynabeads Protein G (ThermoFisher) for 30 min under rotary mixing, and then split in 1.5 mg aliquots of proteins for each assay. Each fraction received 5 µg of antibody and the incubation was left overnight at 4°C under rotation. The following antibodies were used for IP: rabbit anti-DDX17 (ProteinTech 19910-1-AP) or a control rabbit IgG (ThermoFisher), and goat anti-DDX5 (Abcam ab10261) or control goat IgG (Santa Cruz). The next day, the different lysate/antibody mixtures were divided, treated or not with benzonase (Merck-Millipore) for 30 min at 37°C, and then incubated with 50 µl Dynabeads Protein G (ThermoFisher) blocked with bovine serum albumin, for 4 h at 4°C under rotation. Bead were then washed 5 times with IP buffer. Elution was performed by boiling for 5 min in SDS-PAGE loading buffer prior to analysis by western-blotting.

RNA extraction and quantitative PCR analyses

Total RNA were isolated using TriPure Isolation Reagent (Roche). For reverse transcription, 0.5–2 µg of purified RNAs were treated with Dnase I (ThermoFisher) and retrotranscribed using Maxima reverse transcriptase (ThermoFisher), as recommended by the supplier. Potential genomic DNA contamination was systematically verified by performing negative RT controls in absence of enzyme, and by including controls with water instead of cDNA in qPCR assays.

For qPCR analyses, the specificity and linear efficiency of all primers (sequences are in Supplementary Table S4) was first verified by establishing a standard expression curve with various amounts of human genomic DNA or cDNA. qPCR reactions were carried out on 0.625 ng cDNA using a LightCycler 480 System (Roche), with the SYBR[®] Premix Ex Taq (Tli RNaseH Plus) (Takara), under thermocycling conditions that were recommended by the manufacturers. Melting curves were controlled to rule out the existence of non-specific products. Relative DNA levels were calculated using the $\Delta\Delta C_t$ method (using the average C_t obtained from technical duplicates or triplicates) and were normalized to the expression of *GAPDH* RNA.

Detection of nascent transcripts and pull-down of specific mRNA

The metabolic labelling and capture of newly synthesized transcripts was performed using the Click-iT[®] Nascent RNA Capture Kit (ThermoFisher), as recommended by the manufacturer. Briefly, 0.2 mM 5-ethynyl uridine (EU) was

added to the cell culture medium 48 h post-transfection and incubated for 2 h at 37°C to label nascent RNAs. Cells were harvested, RNAs were extracted and treated with DNase. EU-labelled-RNAs (1 µg) were biotinylated by incubation for 30 min at room temperature in a mix containing 1× EU buffer, 2 mM CuSO₄, 0.5 mM biotin azide, 10 mM reaction buffer additive 1 and 12 mM reaction buffer additive 2. RNAs were ethanol-precipitated, resuspended in RNase free water and quantified with a NanoDrop 2000 (ThermoFisher). 200 ng of biotinylated-RNAs were denatured for 5 min at 70°C in 1× RNA binding buffer and RNase OUT Recombinant Ribonuclease Inhibitor, before being immobilized on Dynabeads MyOne Streptavidin T1 magnetic beads (15 µl of beads per condition) for 30 min at RT in a Thermomixer[®] (Eppendorf). Beads were washed as recommended and resuspended in 1 volume of wash buffer 2. Immobilized nascent RNAs bound were reverse transcribed on the beads with the Maxima RT (ThermoFisher) and analysed as described above.

For the pull-down of specific mRNAs, Dynabeads MyOne Streptavidin T1 magnetic beads were washed 3 times with 10 volumes of binding buffer (10 mM Tris pH 7.5, 1 M NaCl, 0.5 mM EDTA, 0.05% Tween-20) and resuspended in 2 volumes of binding buffer. Biotinylated oligonucleotides (0.25 µM, an equimolar mix of 2 different oligonucleotides for each mRNA) were incubated with the beads for 30 min at room temperature on a Thermomixer[®] (Eppendorf). Oligonucleotide-associated beads were then washed 3 times with 10 volumes of binding buffer and resuspended in 1 volume of Wash buffer 1 (from the Click-iT[®] Nascent RNA Capture Kit, ThermoFisher). 200 ng of RNAs were incubated with the beads for 30 min at room temperature on a Thermomixer[®]. Pulled-down RNA were recovered and analysed as described above.

RNA-Seq

Stranded RNA libraries were prepared after enrichment of polyA⁺ RNA. High throughput sequencing of 125 bp paired-end reads was carried out on an Illumina HiSeq 2500 platform (Genewiz), generating an average number of 45 millions of matched pairs of reads. Reads were mapped to the human reference genome (hg19) and the number of reads per gene was counted using HTSeq-Count (72), as detailed previously (62). Alternative splicing analyses were performed using the FaRLine pipeline (73), except intron retention that was analysed using rMATS (74). Significant splicing changes were selected if the difference in the percentage of splicing inclusion (PSI) between DDX5/DDX17-depleted cells and control cells was >10% with a *P*-value (corrected for multiple comparisons) <0.05.

The transcriptional read-through was computed from our RNA-seq analysis using previously generated stranded BAM files (*n* = 3). For each terminal exon from the FasterDB database (19 415 genes), we designed a 5 kb region downstream of the exon. When another gene was present within this window, the size was adjusted to fit with the intergenic distance. The region was then divided in successive 50 nt segments, in which the read coverage was computed using SAMTools (75). A segment was considered positive only when the average coverage was >2. The size of the

read-through region was calculated by iterative extension of the segments, until it was interrupted by two consecutive negative segments. An average coverage and standard deviation was then computed. We performed the same coverage analysis on the last 500 nt of the exon (or the entire size of the terminal exon for smaller exons), which were also divided in 50-nt boxes starting from the 3' end of the gene. The percentage of read-through was then computed for each experimental condition as the ratio between the mean coverage in the downstream region to the mean coverage in the last exon. A visual inspection of the BAM files allowed to filter the data to remove unreliable candidates: we removed genes that displayed an average read-through coverage <20 in the siDDX5/17 condition, a maximum value of computed relative read-through <0.1% in any condition, a negative difference in read-through between the control and siDDX5/17 condition, a null coverage value in the terminal exon in any condition (more upstream termination), and weakly expressed genes (basemean < 100).

Chromatin immunoprecipitation

ChIP experiments were carried out as described previously (62), with some modifications in the composition of several buffers, as indicated below. Chromatin was precipitated with 5 µg of mouse anti-Pol II antibody (F-12, sc55492, Santa Cruz) or with 4 µg of rabbit anti-CTCF (AB_2614975, Active Motif), or equivalent amounts of their corresponding IgG Isotype control (ThermoFisher).

Lysis buffer	Shearing buffer	Buffer D (dilution)	Low salt buffer (1 wash)
5 mM HEPES pH 8.0	10 mM Tris pH 8.0	10 mM Tris pH 8.0	20 mM Tris pH 8.0
85 mM KCl	1 mM EDTA	150 mM NaCl	150 mM NaCl
0.5% NP40	0.1% SDS	1 mM EDTA	0.1% SDS
		1% Triton X-100	1% Triton X-100
		0.01% SDS	2 mM EDTA
High salt buffer (1 wash)	Low LiCl buffer (1 wash)	Tris/EDTA (2 washes)	Elution buffer
20 mM Tris pH 8.0	100 mM Tris pH 8.0	10 mM Tris pH 8.0	200 mM NaCl
500 mM NaCl	0.5 M LiCl	1 mM EDTA	0.1 M NaHCO ₃
0.1% SDS	1% NP-40		1% SDS
1% Triton X-100	1% Na-DOC		20 µg Proteinase K
2 mM EDTA			

For each of the 3 replicates of calibrated RNAPII ChIP-seq, we pooled 4 independent IP, each of them performed on 40 µg of chromatin supplemented with 32 ng of *Drosophila* spike-in chromatin (Active Motif #53083), with 4 µg of mouse anti-Pol II F-12 antibody (F-12, sc55492, Santa Cruz) and 2 µg of spike-in antibody (Active Motif #61686). Libraries and sequencing was performed by the GenomEast platform. ChIP samples were purified using Agencourt AMPure XP beads (Beckman Coulter) and quantified with the Qubit (Invitrogen). ChIP-seq libraries were prepared from 10 ng of double-stranded purified DNA using the MicroPlex Library Preparation kit v2 (C05010014, Diagenode, Seraing, Belgium), according to manufacturer's instructions. High throughput sequencing of 50 bp single-end reads was carried out on an Illumina HiSeq 4000 platform.

Analysis of calibrated RNAPII ChIP-seq data

The analysis of sequencing data was based on a previous protocol (76), details of the Nextflow (77) pipeline are available in the following GitHub repository: [https://gitbio.ens-](https://gitbio.ens-lyon.fr/LBMC/Bernard/quantitative-nucleosome-analysis/)

[lyon.fr/LBMC/Bernard/quantitative-nucleosome-analysis/](https://gitbio.ens-lyon.fr/LBMC/Bernard/quantitative-nucleosome-analysis/) and instructions in README_quantitativ-chip.md to reproduce our analysis). The approach is similar to a standard ChIP-seq analysis, except that the presence of a defined amount of exogenous chromatin (from *Drosophila*) in each sample allows a normalization of the experimental samples and thus a more accurate comparison of RNAPII occupancy across samples. After trimming and removing adaptors, reads were then competitively mapped on a concatenation of the two genomes (the reference human genome hg19, and the calibration *drosophila* genome dm6) and assigned to their corresponding genome. Calibration of the read coverage was then performed for each nucleotide position.

Meta-exon and meta-gene analyses were generated using coverage files in BigWig format obtained and normalized as described above. RNAPII coverage at each nucleotide of a given set of exons/genes was divided into a fixed number of segments of varying length depending on the exon/gene size. The mean coverage within each segment was calculated to obtain a coverage vector of equal size for each exon/gene. For meta-exon figures, the coverage of those vectors was normalized at each position and for each BigWig file with the mean coverage in the region directly downstream of the TSS of the corresponding genes (corresponding to the first percent of the complete sequence of the gene), and weighted by the number of exons for each gene. For the meta-gene representation, the coverage of the vectors was normalized at each position and for each BigWig file with the mean coverage in the region directly upstream of the PAS (corresponding to the last percent of the complete sequence of the gene). Then the mean coverage at each vector position was calculated. This procedure was also applied across a 10 kb interval upstream and downstream regions of all sets of exons/genes. Finally, the mean coverage between the three RNAPII ChIP-seq replicates was calculated for control and siDDX5/DDX17 conditions.

The code used to build those figures is available in the following Gitlab repository: https://gitbio.ens-lyon.fr/LBMC/regards/Projects.Analyses/bigwig_visu.

Analysis of CTCF ChIP-seq data

To analyse the relative proximity of DDX5/DDX17 exons to CTCF binding sites, we first generated a BED file containing a merged list of CTCF peaks retrieved from several ENCODE datasets (Supplementary Table S5). We next calculated the genomic distance (negative or positive for upstream and downstream peaks, respectively) between each exon from the FasterDB database (78) and the center of the closest CTCF peak. The exon boundary that minimizes the distance to the nearest CTCF peak was chosen for this computation. We performed a logistic regression analysis to test if the different exons regulated by DDX5/DDX17 are closer to CTCF peaks than control exons. This analysis was carried out on: (i) the first ($n = 740$) and last exons ($n = 744$) of genes presenting a 3' end processing defect upon DDX5/DDX17 depletion, or all other genes ($n = 16\,040/16\,046$ for first/last exons, respectively); (ii) internal exons regulated by DDX5/DDX17 or by SRSF1, which were split in two groups depending on

their stronger skipping or inclusion induced by depletion of those factors (1751 activated and 160 repressed exons for DDX5/DDX17, 1115 activated and 561 repressed exons for SRSF1) or all other exons not regulated by these factors ($n = 179\ 618$); (iii) the first ($n = 1819$) and last ($n = 1821$) exons of genes containing a DDX5/DDX17-dependent internal exon or other internal exons ($n = 14\ 969/14\ 969$ for first/last exons, respectively). Alternative exons regulated by SRSF1 were obtained from 4 GEO datasets (GSE26463, GSE52834, ENCSR094KBY and ENCSR066VOO) as described previously (79). We modeled the proximity to a CTCF peak according to the different groups of exons using the *glm* function, with family = binomial ('logit') in R software (R Core Team (2018), R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna). A CTCF peak was considered as close to an exon if its center is located within an exon or within an interval of 1 base to 2 kb upstream or downstream of the exon. To test the differences between the groups of exons, a Tukey's test was used (with R, *emmeans* function from *emmeans* library).

Density figures were created to represent the genomic distance of the nearest peak located from 1 bp to approximately 100 kb upstream or downstream of: (i) the first and last exons of genes with a 3'-end processing defect upon DDX5/DDX17 depletion, or all other genes; (ii) exons repressed or activated by DDX5/DDX17 and SRSF1 or all other internal exons.

The orientation of the nearest CTCF peaks upstream and downstream of human exons was computed using GimmeMotifs version 0.16.0 (80) with a motif score cutoff of 0.8 and reporting only the best match (options -c 0.8 and -n 1). The MA0139.1 CTCF motif file used in this analysis was downloaded from JASPAR (<https://jaspar.genereg.net/matrix/MA0139.1/>) (81). We could assign an orientation for 74.4% of CTCF peaks. Then, we computed the proportion of convergent, divergent or tandem pairs of CTCF sites which flank looped exons as defined by ChIA-PET analyses (see the corresponding section for more details). For this analysis, only the 6 datasets corresponding to CTCF/Cohesin ChIA-PET were used (weight = 2), and we analysed the CTCF pairs corresponding to: (i) loops between the first and last exon of genes presenting a 3' end processing defect upon DDX5/DDX17 depletion, or their control loops; (ii) loops between a DDX5/DDX17-dependent internal exon and the first exon of its cognate gene, or their control loops (loops involving SRSF1-dependent exons or other internal exons). We pooled all DDX5/DDX17-regulated exons ($n = 52$) and all controls ($n = 1136$) for more clarity.

Analysis of ChIA-PET data

Processed ChIA-PET BED files were retrieved from the GEO database (Supplementary Table S5 for the list of 20 datasets) and were used to determine which exons interact in pairs. To do so, the genomic coordinates of each exon were retrieved from the FasterDB database (78) and were modified to include a 200 bp window on each side of each exon. We used the BEDTools intersect tool (82) to determine which extended exon matches with a PET anchor.

Then we defined a list of extended exons interacting with another exon for each dataset (a weight was attributed to each exon pair based on the number of PETs that report the association), and stored this information in a database. For subsequent analyses, exon pairs found in at least 3 ChIA-PET datasets with a weight of 2 (i.e. 2 PETs supporting the interaction) were recovered from the ChIA-PET database. As small genes do not contain many reliable intragenic interactions, we only considered exons from genes of size ≥ 15 kb. To avoid considering self ligation PETs, only interactions between exons distant from at least 10 kb were kept for the statistical analysis. These parameters reduced the number of useable PET-tags, especially those connecting internal exons, but they make our analysis more significant. A logistic regression analysis was performed to test if the last exons of DDX5/DDX17-dependent genes ($n = 473$) or other last exons (control exons, $n = 11\ 025$) have a different proportion of interactions with the first exon of their corresponding gene. We used the same statistical methods and functions as those described above to analyse the proximity to CTCF binding sites. The same statistical analysis was performed to define the percentage of interactions between the first and last exon of genes containing an internal DDX5/DDX17-dependent exon ($n = 1020$, $n = 971$ for SRSF1-regulated exons and $n = 9098$ for Ctrl exons), and between the regulated internal exons and the first exon of their cognate genes ($n = 1557$, $n = 1388$ for SRSF1-regulated exons and $n = 154\ 380$ for Ctrl exons).

Weblogo analysis

To check if some motifs are enriched around the PAS, we analysed 400 nucleotides-long sequences centered on the 3' end of the last exon of genes presenting a 3' end processing defect upon DDX5/DDX17 depletion ($n = 744$), or all other genes ($n = 16\ 046$), using the MEME v5.4.1 program of the MEME suite (<https://meme-suite.org/meme/>) (83). The following parameters were used: '-mod zoops -n motifs 10 -minw 6 -maxw 20 -brief 200000 -dna -nostatus'. Then the SEA program of the MEME suite was used to check if the MA0139.1 CTCF motif was enriched in those sequences. The same procedure was applied for 200 nucleotides-long sequences centered on the 3'SS ($n = 2256$) and the 5'SS ($n = 2256$) of internal exons regulated by DDX5/DDX17 or other internal exons ($n = 181\ 294$).

RESULTS

Silencing of DDX5 and DDX17 induces various RNA processing defects

To explore the genome-wide impact of DDX5 and DDX17, we deeply sequenced polyA-enriched RNA ($n = 3$) extracted from human neuroblastoma SH-SY5Y cells in which the expression of both RNA helicases was silenced or not (Figure 1A and Supplementary Figure S1A). We observed that in the absence of DDX5/DDX17, many genes exhibited an altered sequencing coverage beyond their annotated 3' end, suggesting a defect in 3' end RNA processing and/or transcription termination (Figure 1B). Based on the quantification of the reads downstream of the polyadenylation site (PAS) of expressed genes, we defined a list of

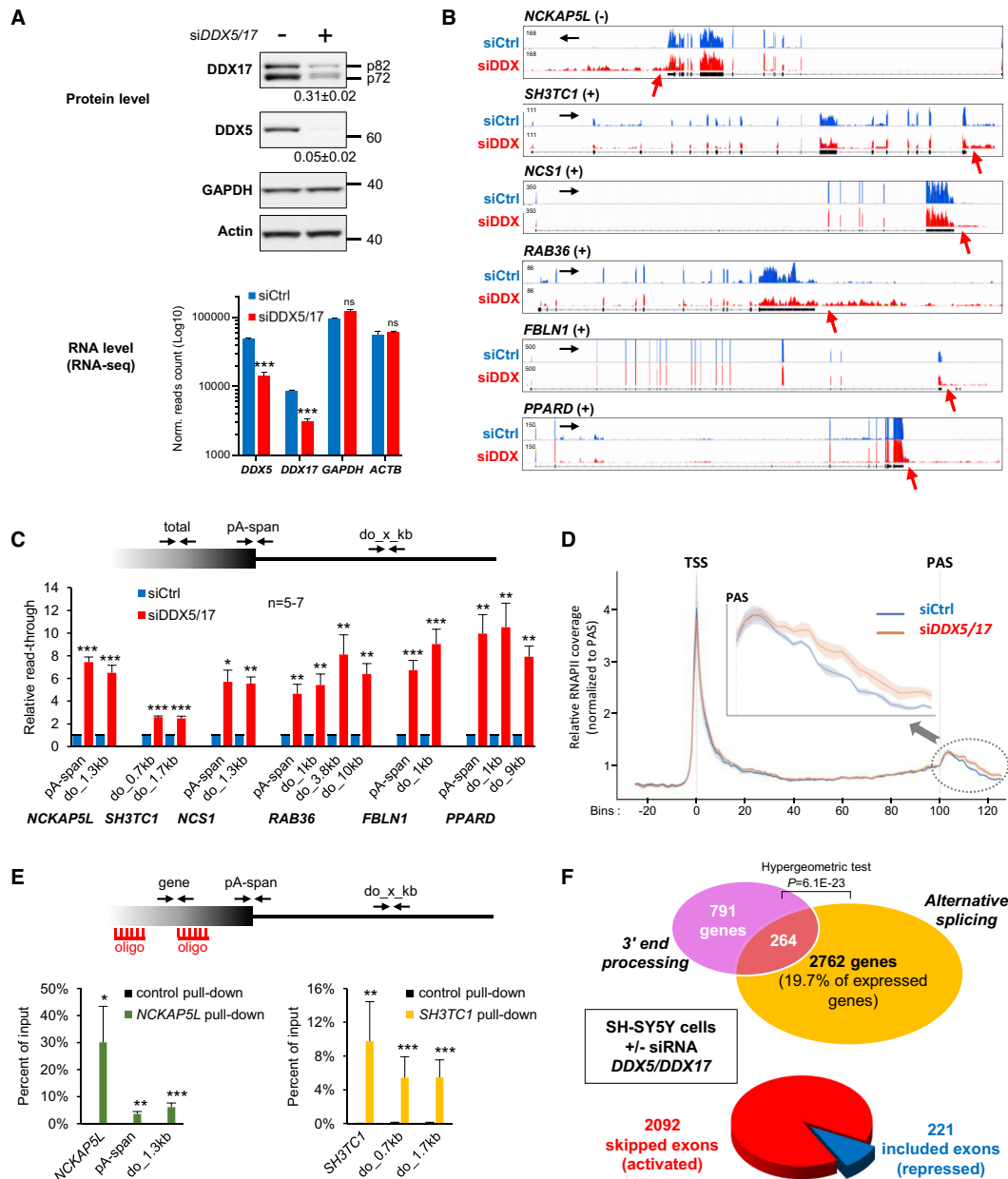


Figure 1. (A) Validation of DDX5 and DDX17 depletion. Quantification of protein level (top) corresponds to the mean expression value normalized to actin \pm S.E.M. ($n = 3$). RNA levels (bottom) were calculated from RNA-seq data and represented as the mean normalized read count \pm S.E.M. ($n = 3$). Unpaired t -test (***) P -val < 0.001. (B) Representative examples of genes presenting an increased RNA-seq coverage beyond their 3' end in condition of silencing of both *DDX5* and *DDX17* (red), compared to a control siRNA (blue). The RefSeq genomic annotation is shown for each gene (in black), as well as the gene orientation (black arrow). The respective width of each window corresponds to 53, 46, 74, 35, 115 and 111 kb. For each gene, all the reads originate from the same strand and there was no antisense transcript. (C) Steady-state quantification of read-through transcripts. The amount of each RNA product, measured by RT-qPCR using primers spanning the PAS (pA-span) or at a variable distance downstream of the gene (do.distance), as indicated on the diagram, was normalized to the total amount of the corresponding regular mRNA, measured with primers localized near the 3' end of the gene (total). Data are represented as the mean values \pm S.E.M. of independent experiments ($n = 5-7$) normalized to the control sample, set to 1. Paired t -test (* P -val < 0.05; ** P -val < 0.01; *** P -val < 0.001). (D) Meta-gene analysis of the distribution of total RNAPII across the genes presenting *siDDX5/DDX17*-dependent transcriptional read-through. The analysis spans from 10 kb upstream of the TSS to 10 kb downstream of the PAS. The close-up view shows only the 10 kb region downstream of the PAS. (E) Quantification of pulled-down transcripts. RNA from *siDDX5/DDX17*-treated cells were pulled-down using biotinylated ASOs targeting constant exonic regions of *NCKAP5L* or *SH3TC1* transcripts. RNA products expressed from these two genes were then quantified in the pulled-down fraction, using primer pairs located in the regular transcript, across the PAS or downstream of the gene, as indicated. For the control pull-down, we used ASOs specific for a different gene (*KATNBI*). Data are represented as the percentage of bound RNA relative to input material (mean value of 3 independent experiments \pm S.E.M.). Paired t -test (* P -val < 0.05; ** P -val < 0.01; *** P -val < 0.001). (F) Venn diagram showing the number of genes presenting a 3' end cleavage defect or at least one alternative splicing event in absence of DDX5/DDX17, as predicted from the RNA-seq data (set with $[\Delta\text{PSI}] > 0.1$). The lower diagram shows the number of alternative cassette exons misregulated upon DDX5/DDX17 depletion. The term 'activated' or 'repressed' refers to DDX5/DDX17 activity on the corresponding exons, *i.e.* exons that are skipped or included upon *siDDX5/DDX17* treatment, respectively.

791 genes that presented this profile compared to the control condition (Supplementary Table S1).

We quantified experimentally the amount of RNA downstream of the expected PAS of 11 randomly selected transcripts (Figure 1C and Supplementary Figure S1B). We measured a relative 2-to-10-fold increase of the level of the extended transcripts upon *DDX5/DDX17* silencing, compared to the control condition and after normalizing to the total steady-state mRNA level. This normalization was required since the steady-state expression of these genes was often reduced upon *DDX5/DDX17* silencing (Supplementary Figure S1C), suggesting a higher instability and/or a reduced transcription. This tendency was confirmed for the whole set of 3' end-impacted transcripts (Supplementary Figure S1D).

We next carried out several experiments to show that these extended mRNA indeed resulted from a weakened RNA cleavage at the annotated polyadenylation site (PAS) and/or altered transcription termination. First, we analysed the occupancy of total RNA polymerase II (RNAPII) across genes via a calibrated ChIP-seq analysis of control and *DDX5/DDX17*-depleted cells. A meta-gene analysis showed an increase of RNAPII density downstream of the polyadenylation site (PAS) of genes whose termination is altered in absence of *DDX5/DDX17* (Figure 1D). Second, the quantification of metabolically labelled nascent transcripts also showed an increase in nascent transcription downstream of the annotated PAS (Supplementary Figure S1E). Third, the pull-down of *NCKAP5L* or *SH3TC1* transcripts using antisense oligonucleotides targeting their last annotated exon specifically recovered the downstream region, indicative of a deficient cleavage of the pre-mRNA after the PAS (Figure 1E). Altogether, these results showed that *DDX5/DDX17* depletion impaired the 3' end processing of nascent transcripts and/or transcription termination of a large group of genes. Of note, the consensus sequence of the PAS of *DDX5/DDX17*-impacted genes matched the usual AAUAAA consensus although the score of the second position was slightly reduced compared to control genes (Supplementary Figure S1F). We also looked for enriched motifs around the PAS of these exons, but apart from a higher frequency of G/C-rich sequences, that are also found around control exons, we did not identify a unique motif that would clearly differentiate them from unregulated exons (Supplementary Figure S1G).

We also quantified splicing variations in our RNA-seq dataset and found important changes upon *DDX5/DDX17* silencing, affecting nearly 20% of expressed genes (Figure 1F, Supplementary Table S2). Focusing on simple cassette exons, we identified more than 2300 *DDX5/DDX17*-dependent exons, a large majority of which were more skipped in absence of both helicases compared to the control condition (Figure 1F, Supplementary Table S2), consistent with our previous studies (60,73). Analysing the expression of 226 splicing regulators did not reveal any significant impact of *DDX5/DDX17* depletion, reducing the possibility of indirect effects (Supplementary Figure S1I, Supplementary Table S3). Besides, we did not find any enriched sequence around the splice sites of *DDX5/DDX17*-regulated exons that could be associated to a known splicing regulator, but this analysis again indicated a pre-

dominance of G- and C-rich sequences (Supplementary Figure S1H).

Interestingly, we observed that 34% of genes that exhibited defective PAS usage upon *DDX5/DDX17* silencing were also impacted at the splicing level, which was significantly higher than randomly expected (Figure 1F). The overlap was still significant when considering only the most significant splicing variations (Supplementary Figure S1J). This suggested that the transcriptional read-through observed in the absence of *DDX5/DDX17* may be linked to a global deregulation of co-transcriptional RNA processing for this subset of genes.

DDX5/DDX17-dependent terminal and internal exons are close to CTCF-binding sites

Searching for a possible mechanism that could link *DDX5/DDX17*, alternative splicing, 3' end cleavage and transcription termination, we tested a possible interaction between both helicases and the Cleavage Factor CFIm, but we did not detect any signal in our co-immunoprecipitation experiments (Supplementary Figure S2A). We then focused our attention on the transcription and chromatin architecture factor CTCF, a known partner of both *DDX5* and *DDX17* (68,69). Indeed, CTCF co-precipitated with endogenous *DDX5* and *DDX17* proteins from SH-SY5Y cell extracts, and this interaction was enhanced upon degradation of nucleic acids by benzonase (Supplementary Figure S2B). It has to be underlined that *DDX5* and *DDX17* are found in both the cytoplasm and the nucleus, so only their nuclear pool could pull-down with CTCF which is strictly nuclear, making this result more significant. Moreover, the consensus CTCF binding motif is enriched around the PAS and splice sites regulated by *DDX5/DDX17* (Supplementary Figure S2C).

Using public CTCF ChIP-seq datasets, we counted the CTCF binding sites localized within *DDX5/DDX17*-regulated genes and found their average number was significantly higher than in other genes (Supplementary Figure S2D). We next built a statistical model to test the probability of finding one CTCF binding site overlapping with, or within 2 kb of, *DDX5/DDX17*-dependent exons. We found a strong enrichment of CTCF binding at the level of the last exon (in purple) of *DDX5/DDX17*-regulated genes, compared to unaffected genes (Figure 2A). A highly significant enrichment was also observed 2 kb downstream, and to a lesser extent 2 kb upstream, of those terminal exons. Moreover, our analysis showed that the first exon of *DDX5/DDX17*-regulated genes was also frequently associated to CTCF binding (with a clear bias upstream of the first exon). Plotting the distance between exon boundaries and the closest encountered CTCF binding site also confirmed the higher proximity of *DDX5/DDX17*-regulated genes to CTCF sites, on both sides of their first and last exons (Supplementary Figure S2E). These data suggested that CTCF often binds on both sides of those genes.

We next analysed internal alternative exons (Figure 2B and Supplementary Figure S2F), considering *siDDX5/DDX17*-induced skipped (red, top panel) and included (blue, bottom panel) exons separately as they have different intrinsic properties (60). As controls, we used ex-

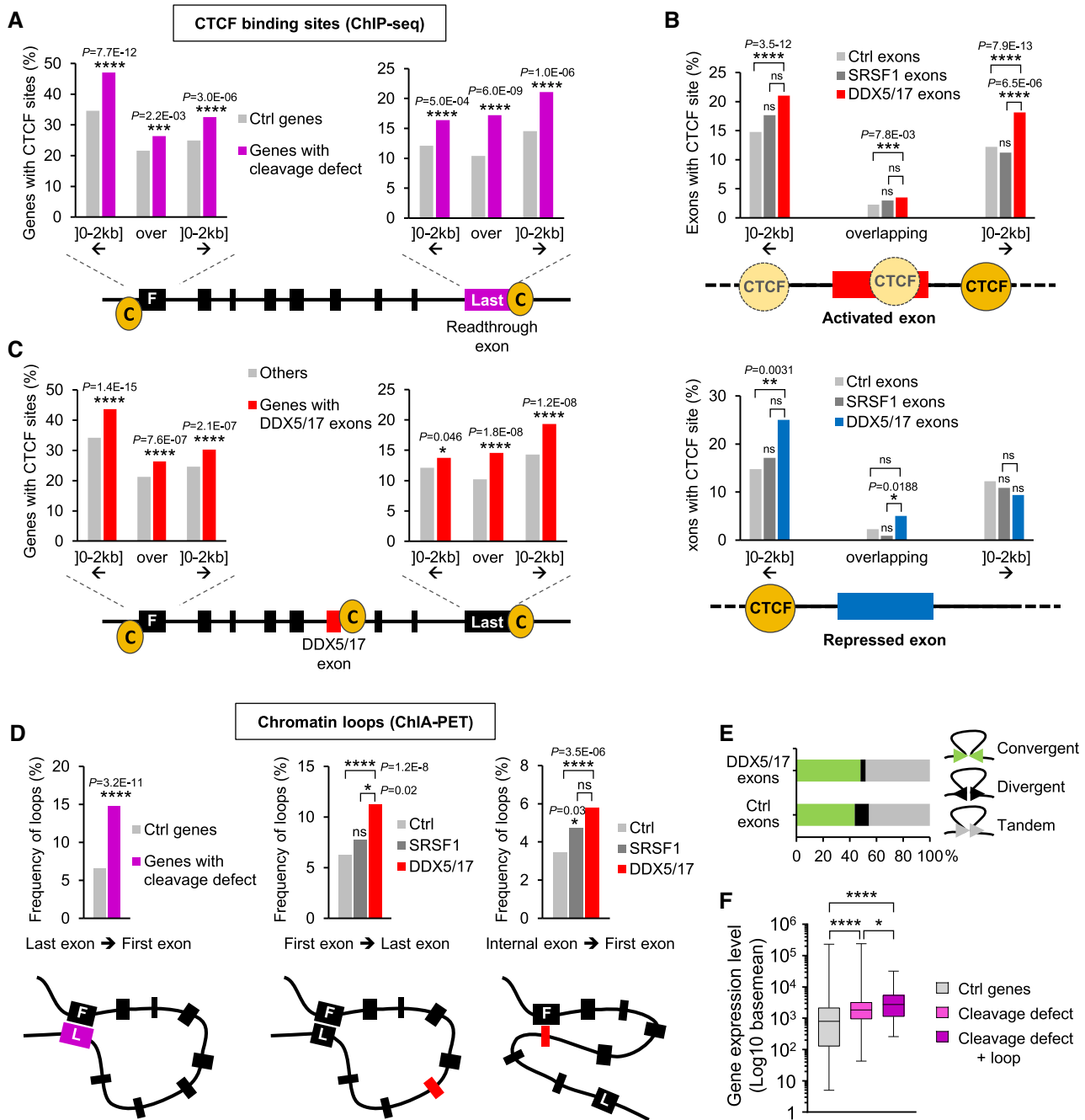


Figure 2. (A) Relative distribution of CTCF binding sites at and around the first (F) and last exons of genes showing *siDDX5/DDX17*-induced defect in 3' end cleavage (purple) compared to unregulated genes (Ctrl, grey). (B) Relative distribution of CTCF binding sites at and around the exons that are skipped (red, top panel) or included (blue, bottom panel) upon *DDX5/DDX17* depletion. Control sets of exons include exons skipped or included upon depletion of the *SRSF1* splicing regulator, and all internal exons that are neither dependent on *DDX5/DDX17*, nor dependent on *SRSF1* (Ctrl exons). (C) Relative distribution of CTCF binding sites at and around the first (F) and last exons of genes containing at least one internal *DDX5/DDX17*-dependent alternative exon (in red). (D) Frequency of intragenic looping in genes harboring *DDX5/DDX17*-dependent exons, based on ChIA-PET datasets (for CTCF, Cohesin and RNAPII). Left panel: looping between the first and last exon of genes with *siDDX5/DDX17*-induced defect in 3' end cleavage (purple), compared to unregulated genes. Central panel: looping between the first and last exons of genes containing *DDX5/DDX17*-dependent alternative exons (red), *SRSF1*-dependent exons or other exons. Right panel: looping between the internal exon regulated by *DDX5/DDX17* or *SRSF1* and the first exon of its cognate gene. (E) Orientation of the pairs of CTCF sites corresponding to chromatin loops analysed in D between the first and internal or last exons. Only exon pairs found in ChIA-PETs datasets using CTCF, SMC1 or RAD21 antibodies were selected for this analysis, with a weight 2 in at least one dataset. The different orientations of CTCF site pairs are depicted on the right. (F) Basal steady-state expression of transcripts of which 3' end cleavage is regulated (light purple) or not (grey) by *DDX5/DDX17*. Within *DDX5/DDX17*-dependent genes, those that present evidence for head-to-tail proximity (dark purple) are even more highly expressed than other genes. Only transcripts with a basemean expression >5 were considered for the analysis. ANOVA with Kruskal-Wallis non-parametric tests (* *P*-val < 0.05; **** *P*-val < 0.0001).

ons deregulated upon depletion of the serine/arginine-rich splicing factor SRSF1, as well as all internal exons that were regulated neither by DDX5/DDX17 nor by SRSF1 (Ctrl exons). Again, there is a significant enrichment of CTCF binding at and around DDX5/DDX17-activated exons (i.e. skipped upon *siDDX5/DDX17* treatment), with a more visible bias for the region downstream of the exon, compared to SRSF1 or Ctrl exons (Figure 2B, top panel, and Supplementary Figure S2F, left panel). This is consistent with the observation that CTCF is more likely to regulate exon inclusion when it binds downstream of the exon (45). Extending this analysis to the first and last exons of genes containing an alternative DDX5/DDX17-dependent exon, we found a significant enrichment of CTCF binding at their both extremities compared to control genes, as for 3' end-regulated genes (compare Figures 2A and C).

In contrast, DDX5/DDX17-repressed exons (i.e. included upon *siDDX5/DDX17* treatment) exhibited a different pattern: only a modest CTCF enrichment was detected within 2 kb upstream of these exons compared to Ctrl exons, and no enrichment was found over or downstream of the exons (Figure 2B, bottom panel, and Supplementary Figure S2F, right panel). As we examined further this subset of repressed exons, we discovered that their intragenic position was strongly biased toward the 5' end of genes, with nearly 40% of DDX5/DDX17-repressed exons annotated in position 2 or 3, which is twice the number as the more evenly distributed DDX5/DDX17-activated exons (Supplementary Figure S2G). This promoter-proximal bias probably explains the binding of CTCF upstream of these exons, as many CTCF binding sites are found within 2 kb of the TSS (84). DDX5/DDX17-repressed exons are also characterized by a low basal inclusion level (60), and many of these exons were barely included in the control condition. In fact, we identified several unannotated promoter-proximal exons, for example in the *RARS* and *SGTA* genes, that were included only upon depletion of DDX5/DDX17 (Supplementary Figure S2H).

Altogether these results revealed a specific pattern of CTCF binding on genes regulated by DDX5/DDX17, suggesting a relationship between the binding of CTCF and DDX5/DDX17-dependent RNA processing of internal and terminal exons.

Frequent intragenic looping in DDX5/DDX17-regulated genes

We next asked whether CTCF binding at different intragenic locations along DDX5/DDX17-regulated genes could reflect the existence of chromatin contacts between those specific sites.

To address this question, we used existing ChIA-PET (Chromatin Interaction Analysis by Paired-End Tag sequencing) datasets, which identified the DNA fragments that are brought together in close spatial proximity by a given factor. In agreement with CTCF ChIP-seq data (Figures 2A and C), we found a significantly higher proportion of PET-tags connecting the first and last exon of genes whose 3' end processing was altered upon DDX5/DDX17 depletion than for unaffected genes (Figure 2D, left panel). Similarly, we observed a higher frequency of contacts be-

tween the first and last exon within the group of genes containing a DDX5/DDX17-dependent alternative exon, and not in those containing a SRSF1-dependent exon (Figure 2D, middle panel). Finally, we found a higher proportion of connections involving DDX5/DDX17-dependent exons and the first exon of their gene, compared to control internal exons (Figure 2D, right panel). SRSF1-regulated exons are also frequently connected to their cognate promoter, albeit to a lesser extent, suggesting that such intragenic looping may be common for alternative internal exons, as suggested earlier (51–54). We checked the orientation of the pairs of CTCF binding sites that flank the loops connecting internal or terminal exons, and found an equal proportion of convergent and tandem-orientated sites (Figure 2E).

Altogether, these results indicate that a significant subset of DDX5/DDX17-dependent genes are spatially organized with intragenic chromatin loops connecting positions such as their promoter, an internal alternative exon and their 3' terminal region. They also suggest that those loops may be functionally related to the expression and processing of their transcripts. In agreement with a role of gene looping in sustaining efficient gene expression (40), genes whose 3' end cleavage is dependent on DDX5/DDX17, and in particular genes presenting a first-to-last exon looping pattern, tend to be more expressed than unregulated genes (Figure 2F and Supplementary Table S3).

A subset of DDX5/DDX17-dependent exons is coregulated by CTCF

We next tested the contribution of CTCF to 3' end processing and splicing of DDX5/DDX17-regulated transcripts, by reducing its expression in SH-SY5Y cells, in absence or presence of DDX5/DDX17. CTCF depletion did not affect the expression of the two helicases, and vice-versa (Figure 3A).

By itself, *CTCF* knockdown had only a moderate effect on some of DDX5/DDX17-regulated genes, either on 3' end cleavage (*NCSI*, Figure 3B) or on the inclusion of alternative exons (in the *ABLIM1*, *PAM* or *PRMT2* genes, Figure 3C). This is in line with previous studies that reported only minor effects on transcription, even under conditions of acute CTCF depletion (85–90). Nevertheless, when combined to the silencing of both helicases, the reduction of CTCF level consistently amplified the 3' end defect observed on most genes, compared to the effect of *siDDX5/DDX17* alone (Figure 3B, compare red and black bars). A similar effect was observed for alternative exons that were close to a previously identified CTCF binding site, but not for *URBI* exon 23 and *NCSI* exon 8 that we used as controls because the nearest CTCF site was more than 5 kb away (Figure 3C, compare red and black bars, Supplementary Figure S3A). Together with the analyses of Figure 2, these results supported the idea that CTCF and DDX5/DDX17 are cooperatively involved in the regulation of alternative splicing, 3' end RNA processing and transcription termination of a subset of genes.

To address more directly the role of CTCF in RNA processing, we generated a stable SH-SY5Y cell line in which the CTCF binding site located 1.2 kb downstream of the PAS of the *SH3TC1* gene, within the transcriptional read-

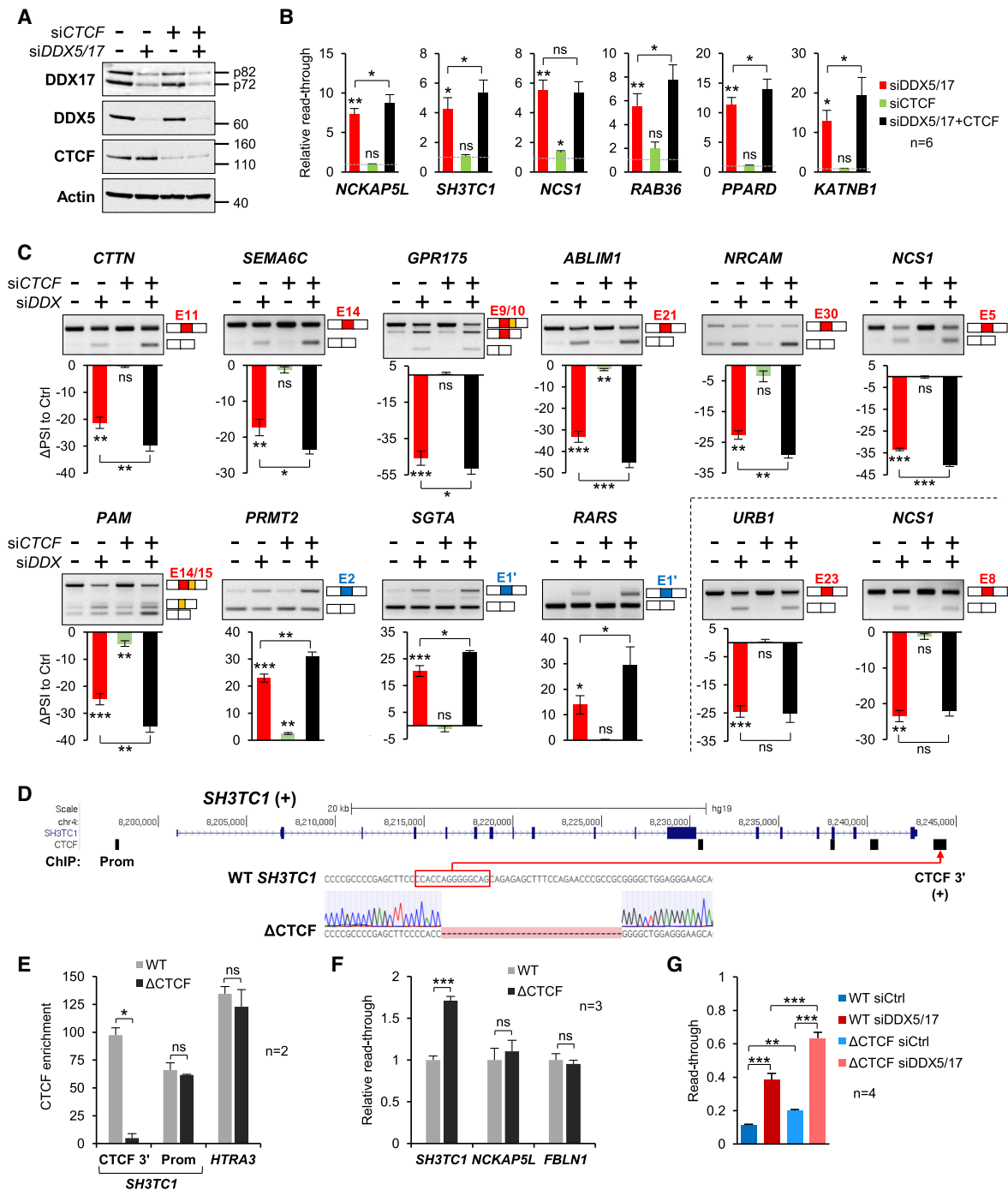


Figure 3. (A) Western-blot showing the expression of DDX5, DDX17 and CTCF in presence of siRNA targeted against *DDX5/DDX17* and *CTCF* transcripts. (B) Quantification of the transcriptional read-through induced by DDX5/DDX17 and/or CTCF depletion on selected genes. Details are as in Figure 1B. Data are represented as the mean value \pm S.E.M. of independent experiments ($n = 6$). Statistical comparison between each condition (including the unshown control condition) was calculated using a one-way ANOVA (Holm–Sidak’s multiple comparison tests: * P -val < 0.05; ** P -val < 0.01; *** P -val < 0.001). (C) RT-PCR analysis measuring the inclusion of a selection of alternative exons in absence of DDX5/DDX17 and/or CTCF. The Δ PSI corresponds to the difference between the PSI (percent spliced-in) score of each depleted sample and the control sample. Details are as in (B). (D) Genomic organization of the *SH3TC1* gene, with the position of CTCF binding sites. The red arrow indicates the deleted CTCF site (CTCF 3’), which is in the same orientation as the gene (+). The bottom panel shows the sequence of the region around the CTCF site (boxed in red), and the resulting sequence in the Δ CTCF cell line. (E) ChIP-qPCR analysis in the parental (WT) and Δ CTCF cell lines. Data are represented as the mean binding enrichment of CTCF compared to a negative gene-free region \pm S.E.M. of two independent experiments. (t -test, * P -val < 0.05). (F) Quantification of the basal level of read-through transcripts in the parental (WT) and Δ CTCF cell lines. Data were normalized to the expression of total transcripts, as described in Figure 1, and then expressed as the mean reported to the read-through observed in WT cells, set to 1 (t -test, *** P -val < 0.001). (G) Quantification of *SH3TC1* read-through transcripts in WT and Δ CTCF cells, in presence (siCtrl) or in absence (siDDX5/17) of DDX5/DDX17. For each condition, data are expressed as the mean value \pm S.E.M. of the amount of read-through products (measured with the ‘do.1.7kb’ primers) normalized to total *SH3TC1* transcripts ($n = 4$). Two-way ANOVA corrected for multiple comparisons by controlling the false discovery rate (** q -val < 0.01; *** q -val < 0.001).

through region, was deleted using a CRISPR-Cas9 approach (Figure 3D). This deletion reduced CTCF binding to background level specifically at this site (Figure 3E). In normal culture conditions, these cells displayed a 1.7-fold increase in the amount of read-through *SH3TC1* transcripts compared to the parental cell line, whereas other genes remained unaffected (Figure 3F). This indicated that CTCF binding indeed contributes to correct 3' end processing and transcription termination. Yet, RNA helicases are still required for this process in the modified cell line, since the depletion of DDX5/DDX17 enhanced transcriptional read-through even further, to a significantly higher level compared to depleted parental cells (Figure 3G and Supplementary Figure S3B).

DDX5/DDX17 regulate CTCF binding and 3D gene organization

We then sought to test the link between CTCF binding and chromatin looping on DDX5/DDX17-dependent model genes. We first focused on the *NCSI* gene, whose 3' end processing is altered upon DDX5/DDX17 depletion (Figure 1), and which contains two internal DDX5/DDX17-dependent alternative exons: exon 5 (E5) is immediately upstream of 2 CTCF-binding sites and its inclusion is altered upon CTCF depletion, whereas exon 8 (E8) is not flanked by any reported CTCF site and is included in a CTCF-independent manner (Figures 3C and 4A).

We analysed the binding of CTCF along the *NCSI* gene in SH-SY5Y cells (Figure 4A and B), and detected strong binding at the terminal exon (E9) and downstream of the gene (3'), two sites that are highly conserved across human cell lines (Supplementary Figure S4A). CTCF binding was also detected at the promoter and immediately downstream of E5 (I6), whereas only background binding was detected within intron 1. Importantly, DDX5/DDX17 depletion significantly reduced CTCF occupancy at the promoter, at the level of E5 and around the 3' end of the gene (Figure 4B). These results suggested that DDX5/DDX17-dependent regulation of *NCSI* E5 inclusion, 3' end RNA cleavage and transcription termination may be related to the binding of CTCF to chromatin.

To analyse the 3D architecture of the *NCSI* gene, we then performed chromosome conformation capture (3C) assays, which quantify the ligation efficiency between restriction fragments maintained in close spatial proximity by cross-linking. We carried out walking quantitative PCR using an anchor primer in the *NCSI* promoter and different primers in downstream restriction fragments along the gene, distant from 5 to 70 kb (Figure 4A, in blue). If the tested chromatin region is unstructured, the ligation efficiency (and the following amplification) between two restriction fragments is expected to decrease as the distance from the anchor increases, while an increased signal in a distant region reveals a physical proximity with the anchor. Our analysis (Figure 4C) showed a progressive decrease of the signal (fragments F1 to F6), only interrupted by a small peak at the level of exon 5 (F4). The signal increased again at the level and downstream of terminal exon 9 (F7 and F8). This suggested the *NCSI* promoter could make some contact with E5, but most evidently that it is spatially close to

the 3' end of the gene (Figure 4C), in line with the location of DDX5/DDX17-dependent CTCF binding sites and ChIA-PET analyses (Figure 2 and Supplementary Figures S4A and S4B). Importantly, this loop between the promoter and the 3' end of the gene was significantly reduced upon depletion of DDX5/DDX17 (Figure 4C, compare the blue and red curves at F7 and F8).

Altogether these data suggested that DDX5/DDX17-dependent chromatin looping across the *NCSI* gene, at CTCF binding sites, is linked to the transcriptional regulation of this gene and to the processing of its transcripts (Figure 4D).

We next analysed in a similar manner the *PRMT2* gene, whose promoter-proximal exon 2 is more included upon knockdown of both DDX5/DDX17 and CTCF (Figure 3C). Our 3C experiments exposed a chromatin contact between the *PRMT2* promoter and the 3' end of the gene (Figure 4E and F, blue curve). As for the *NCSI* gene, DDX5/DDX17 silencing destabilized this gene loop (Figure 4F, red curve) and reduced CTCF binding at the 3' end of the gene (Figure 4G). In these conditions, the efficiency of 3' end cleavage of *PRMT2* transcripts was reduced (Figure 4H), suggesting that the DDX5/DDX17-dependent binding of CTCF links chromatin looping to *PRMT2* RNA processing.

One possibility is that the inclusion of *PRMT2* exon 2 could be promoted by an altered kinetics of the transcription process, possibly associated to the conformational change of the gene (Figure 4I). We tested this hypothesis in two different ways. First, we monitored RNAPII occupancy within the region surrounding the DDX5/DDX17-repressed exons of *RARS*, *CHCHD4* and *PRMT2* genes. Upon depletion of DDX5/DDX17, RNAPII was enriched within the first intron and around the regulated exon of the three genes, compared to the control condition (Supplementary Figure S4C). We next reasoned that a chemically induced reduction of transcription efficiency could have a similar effect as DDX5/DDX17 depletion, and we treated cells with flavopiridol (FP) or 5,6-dichloro-1- β -D-ribofuranosylbenzimidazole (DRB). These two inhibitors target the CDK9 kinase subunit of the positive transcription elongation factor B (P-TEFb), which is involved in the early transition that engages the RNAPII complex into productive elongation. Indeed, both FP and DRB mimicked the effect of DDX5/DDX17 silencing and induced the inclusion of promoter-proximal exons (Supplementary Figure S4D). These results suggest that the inclusion of this particular subset of poorly recognized promoter-proximal exons is kinetically favoured by a change in RNAPII dynamics induced by DDX5/DDX17 depletion.

CTCF-associated exons are associated with a high density of RNA polymerase II

The binding of CTCF near weak alternative exons was proposed to act as a roadblock for RNAPII, which promotes exon inclusion due to a more favourable kinetics of splicing (45). Similarly, CTCF/Cohesin binding between two alternative PAS was recently shown to promote the use of the proximal PAS, which was interpreted as a physical block to RNAPII (50).

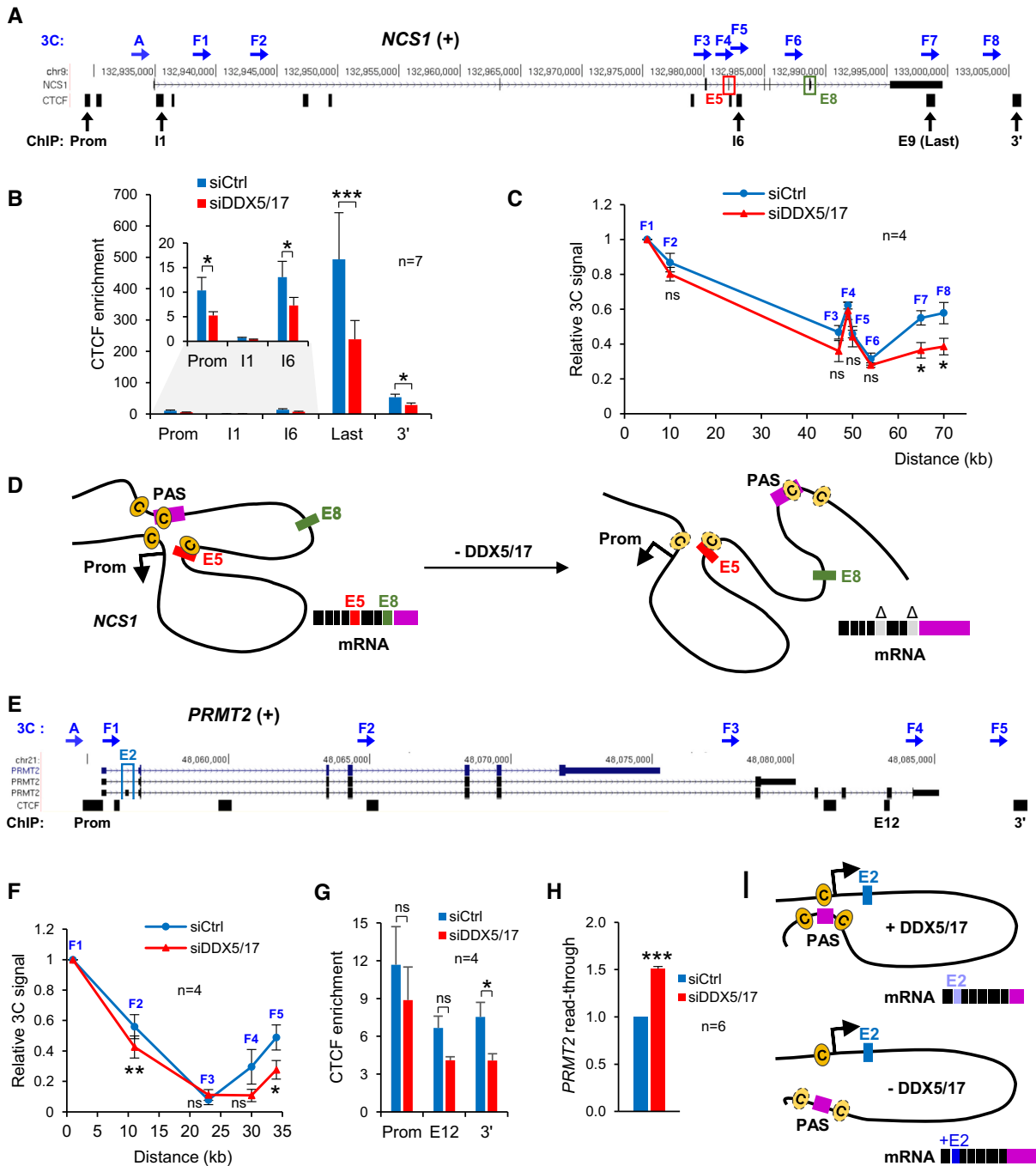


Figure 4. (A) Genomic organization of the *NCS1* gene, with the position of CTCF binding sites. Exon 5 (co-regulated by DDX5/DDX17 and CTCF) and exon 8 (regulated only by DDX5/DDX17) are framed in red and green, respectively. The position of primers used for ChIP-qPCR (black) and 3C (blue) experiments is indicated. (B) ChIP-qPCR analysis showing the effect of DDX5/DDX17 depletion on CTCF binding at various positions along the *NCS1* gene. Data are represented as the mean binding enrichment of CTCF compared to a negative gene-free region, \pm S.E.M. of independent experiments ($n = 7$). Paired *t*-test (* P -val < 0.05; *** P -val < 0.001). (C) 3C experiment showing the relative spatial proximity between various sites of the *NCS1* gene and the promoter (anchor, A), in presence or absence of DDX5/DDX17. The X-axis represents the distance (in kilobases) of each primer relative to the Anchor. Data are represented as the mean signal normalized to the signal at the anchor \pm S.E.M. ($n = 4$ independent experiments). Mann-Whitney test (* P -val < 0.05). (D) Proposed folding of the *NCS1* gene around CTCF sites (orange circles). Only the regulated exons are represented. Upon DDX5/DDX17 depletion, the 3D organization of the gene is altered, especially contacts between the promoter and the 3' end region. At the RNA level this is associated with altered splicing and 3' end cleavage. (E) Genomic organization of the *PRMT2* gene, with the position of CTCF binding sites. Promoter-proximal exon 2 is framed in blue. The position of primers used for ChIP-qPCR (black) and 3C (blue) experiments is indicated. (F) 3C experiment showing the relative spatial proximity between various sites of the *PRMT2* gene and the promoter (anchor, A), in presence or absence of DDX5/DDX17. Details are as in (C). (G) ChIP-qPCR analysis showing the effect of DDX5/DDX17 depletion on CTCF binding along the *PRMT2* gene. Details are as in (B). (H) Quantification of the transcriptional read-through induced by DDX5/DDX17 depletion on the *PRMT2* gene. Details are as in Figure 1B. (I) Proposed folding of the *PRMT2* gene around CTCF sites (orange circles). Upon DDX5/DDX17 depletion, the contact between the promoter and the 3' end of the gene is altered. At the RNA level, this is associated with increased exon 2 inclusion and altered 3' end cleavage.

To further evaluate the functional link between DDX5/DDX17 and CTCF occupancy at regulated exons, we assessed the impact of DDX5/DDX17 on RNAPII distribution, focusing first on the *NCSI* gene analysed by ChIP-qPCR. In both conditions, we normalized the amount of RNAPII to its level near the TSS, to take the difference in gene expression into account. We observed that near the CTCF binding sites of both exons 5 and 9 (last exon), the RNAPII level was comparable to the amount engaged at the gene promoter, and much higher than at control exons 7 and 8 (Figure 5A, blue graphs), suggesting it accumulated in those regulated regions. DDX5/DDX17 depletion did not significantly change RNAPII density, although we noticed a trend toward an increase at all exons (Figure 5A, red graphs). However, since CTCF binding is reduced at both exons 5 and 9 in absence of DDX5/DDX17 (Figure 4B), the high RNAPII density cannot result only from a mere CTCF-dependent roadblock.

To extend these results, we carried out a meta-exon analysis of our calibrated RNAPII ChIP-seq experiments, splitting DDX5/DDX17-activated exons in two groups depending on their position near or far from a CTCF binding site. Remarkably, while RNAPII density remained relatively stable over and around exons that are distant from a CTCF site, it appeared much higher at CTCF-associated exons (Figure 5B). The increase in RNAPII density started in a near 1 kb window upstream of these exons, a region that exhibits a rather low RNAPII density in the other group of exons. The density remained elevated downstream of CTCF-associated exons, where CTCF binding sites are enriched (see Figure 2B). Moreover, depletion of DDX5/DDX17 tended to increase RNAPII density at CTCF-associated exons, although the difference was not statistically significant, while it remained constant at other DDX5/DDX17-dependent exons (Figure 5B, compare red and blue curves).

We analysed in the same way the last exons of genes that exhibit a deficient 3' end cleavage upon *DDX5/DDX17* silencing, and found that CTCF-associated exons similarly displayed a markedly higher density of RNAPII than other exons (Figure 5C). Both unaffected terminal exons (left panel) and DDX5/DDX17-dependent exons distant from CTCF sites (middle panel) had a very similar RNAPII profile, except that the second group lacked a small peak of RNAPII across the 5' boundary of exons. In contrast, RNAPII density at DDX5/DDX17-dependent, CTCF-associated exons (right panel) was higher in a 1 kb window at the 3' end of their upstream flanking region, and it remained higher across the exon. The peak downstream of the PAS, where RNAPII is expected to pause to promote the proper cleavage of the transcript, was visible in all groups but was particularly prominent for CTCF-associated exons (Figure 5C). Note that DDX5/DDX17 depletion did not strongly impact RNAPII density, except downstream of CTCF-associated terminal exons, where transcriptional read-through occurs (as in Figure 1C). Therefore, results of Figures 5A–C indicate that high RNAPII density distinguishes DDX5/DDX17-dependent exons that are associated to CTCF binding, whatever their position within the gene. This feature is unlikely to be a consequence of an increased transcription, as the impact of DDX5/DDX17 de-

pletion is similar on all genes containing DDX5/DDX17-dependent exons, whether these exons are close to a CTCF site or not (Supplementary Figures S5A and S5B).

As evoked above, the high RNAPII density at CTCF-associated exons in absence of DDX5/DDX17 does not fit with the simple model of a roadblock imposed by CTCF to allow exon or PAS recognition. However, DDX5/DDX17 are also required at the RNA level to unwind secondary structures that inhibit the recognition of splicing signals (58,61,62). In fact, DDX5/DDX17-regulated exons and their immediate environment are generally characterized by a high GC content (60,79), which is known to affect transcription (91–93). We re-analysed this parameter for the groups of exons defined above, and found that CTCF-associated exons (and their flanking intronic regions) have a significantly higher GC content than other DDX5/DDX17-activated exons (Figure 5D). Similarly, DDX5/DDX17-dependent terminal exons have a significantly higher GC content than unregulated exons, and it is further increased for CTCF-associated exons (Figure 5E).

We conclude that the GC-rich environment around CTCF-associated exons is associated with an increased RNAPII density. This GC richness could promote the formation of RNA secondary structures that block splicing or 3' end cleavage in absence of DDX5/DDX17. Such structures, involving RNA and/or DNA, may also in some way disturb the progression of RNAPII through this region, especially in absence of DDX5/DDX17.

DNA looping modulates alternative splicing and polyadenylation

Our results and previous reports (50,53) support the idea that several aspects of gene expression and RNA processing can be impacted by intragenic chromatin loops, but direct proof for a link between DNA looping and RNA processing is still lacking. To fill this gap, we adapted the previously described CLOuD9 (chromatin loop reorganization using CRISPR-dCas9) strategy (71) and engineered a stable HEK293 cell line expressing 2 nuclease-deficient dCas9 proteins, fused to protein domains that dimerize only in presence of the phytohormone *S*(+)-abscisic acid (ABA). Upon transfection of standard CRISPR guide RNAs (gRNAs), the addition of ABA to the culture medium allows the *de novo* formation of a loop between the targeted regions.

We tested the CLOuD9 system on an internal PAS in the *FBLN1* gene, to form a loop with its cognate promoter (Figures 6A and B). The 3C analysis confirmed the ABA-induced contact between the two regions (Figure 6C, fragment F4). In these conditions, the amount of transcripts polyadenylated at the PAS1 increased weakly but significantly, while the amount of longer transcripts was concomitantly reduced (Figure 6D, left panel), increasing significantly the ratio between short and long mRNAs (right panel).

We then targeted the alternative exon 7 from the *EYA3* gene (Figure 6E), which combined several features: its basal inclusion level of about 50% allows it to be modulated in both ways, it is not regulated by DDX5/DDX17 and is distant from an identified CTCF binding site. We validated the formation of the loop between the *EYA3* promoter and exon

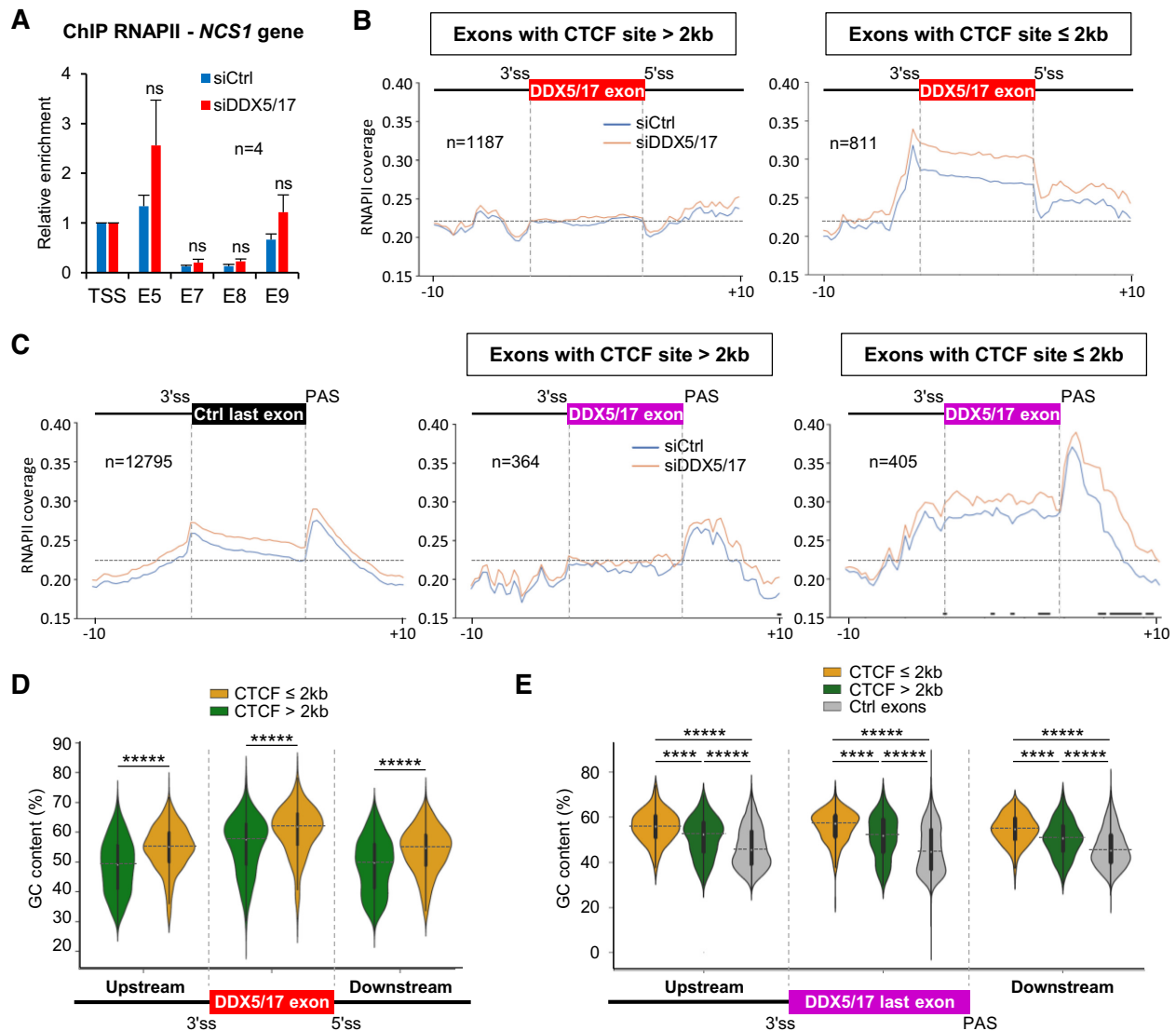


Figure 5. (A) ChIP-qPCR analysis showing the relative binding of RNAPII to several exons of the *NCS1* gene, relative to the beginning of the gene (TSS). The genomic organization of the *NCS1* gene is shown in Figure 4A. Details are as in Figure 4B. (B) Meta-exon analysis of the distribution of RNAPII across DDX5/DDX17-dependent exons. Exons were split into two groups depending on their distance to the closest CTCF binding site (left: exons distant from a CTCF site; right: exons close to a CTCF site). The analysis extended across 10 kb windows upstream and downstream of the exons. For each condition (siCtrl and siDDX5/DDX17) the mean RNAPII coverage ($n = 3$) was normalized to the TSS of the genes. (C) Meta-exon analysis of the distribution of RNAPII across terminal exons. DDX5/DDX17-dependent exons (split into 2 groups as in B) were also compared to other unregulated terminal exons. Details are as in (B). Black lines at the bottom indicate the bins at which a statistical difference in RNAPII coverage was found between the two conditions (paired t -test, P -val < 0.05). (D) GC content of DDX5/DDX17-dependent internal alternative exons (as in B) and their 2 kb intronic flanking regions. Wilcoxon test (**** P -val < $1e-12$; ***** P -val < $1e-16$). (E) GC content of DDX5/DDX17-dependent terminal exons (as in C) and their 2 kb intronic flanking regions.

7 in the presence of ABA, as demonstrated by the increase of the 3C signal at fragment F4 (Figures 6F and G). In these conditions, E7 inclusion was significantly increased (Figure 6H, left panel), while ABA treatment was ineffective on two flanking exons of the *EYA3* gene (middle and right panels), or when non-specific gRNAs were used (Ctrl histograms). In conclusion, these results demonstrate for the first time that creating a chromatin loop between a promoter and an alternative internal or terminal exon is sufficient to stimulate the processing and inclusion of this exon at the RNA level.

DISCUSSION

Here, we show that silencing of RNA helicases DDX5 and DDX17 has a widespread effect on 3' end processing and/or transcription termination, in agreement with previous observations made on a few genes (64,65). DDX5 and DDX17 were proposed to resolve R-loops that form downstream of the PAS in relation with transcription termination, thereby promoting RNAPII release, but whether this could apply to all genes controlled by DDX5/DDX17 is still unknown. Interestingly, in budding yeast the mechanism associated to the transcriptional read-through induced by the loss of

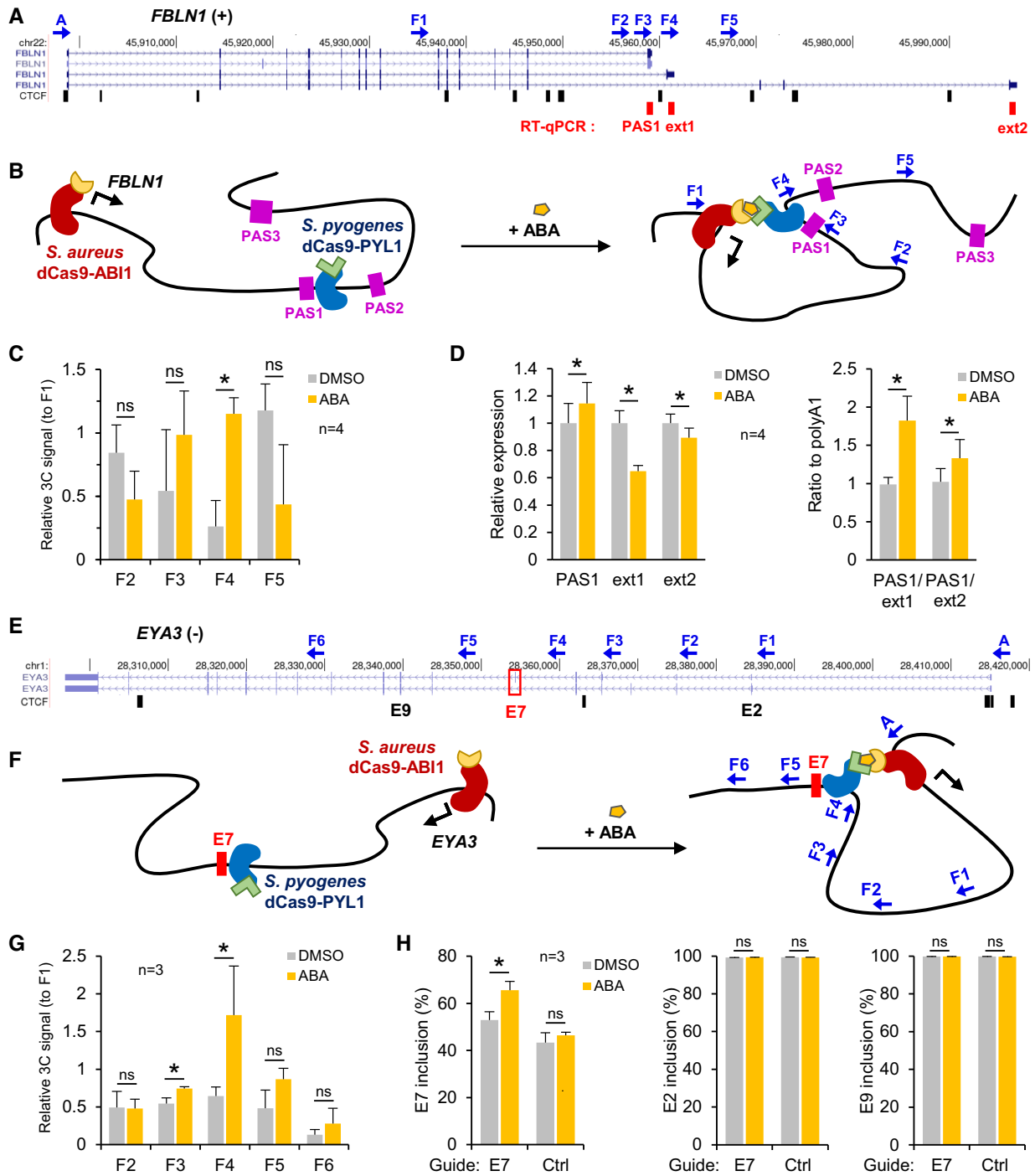


Figure 6. (A) Genomic organization of the *FBLN1* gene, with the position of CTCF binding sites (black), primers used for 3C experiments (blue) and RT-qPCR amplicons (red). (B) CLOuD9 strategy for the *FBLN1* gene. The specific gRNAs targeted the two dCas9 proteins at the promoter and downstream of the proximal PAS (PAS1), allowing the *de novo* formation of a loop between these loci upon addition of abscisic acid (ABA). (C) 3C-PCR experiment (using *DpnII* enzyme) showing the looping between the promoter and the PAS1 of the *FBLN1* gene upon addition of ABA. Data are represented as the mean value \pm S.E.M. of independent experiments ($n = 3$). Mann-Whitney test (* P -val < 0.05). (D) RT-qPCR quantifying the relative amount of transcripts using *FBLN1* PAS1 compared to longer transcripts using PAS2 or PAS3 (ext2 and ext3), in the presence of DMSO or ABA. The right panel represents the data as the ratio between PAS1 and ext2 or ext3 transcripts. Data are represented as the mean value \pm S.E.M. of independent experiments ($n = 4$). Paired t -test (* P -val < 0.05). (E) Genomic organization of the *EYA3* gene, with the position of CTCF binding sites and primers used for 3C experiments (in blue). The alternative exon 7 is framed in red, control exons 2 and 9 are also indicated. (F) CLOuD9 strategy for the *EYA3* gene. The specific gRNAs targeted the two dCas9 proteins respectively at the promoter and near exon 7, allowing the *de novo* formation of a loop between these loci upon addition of abscisic acid (ABA). (G) 3C-PCR experiment (using *HindIII* enzyme) showing the looping between the promoter and exon 7 of the *EYA3* gene upon addition of ABA. Details are as in C. (H) RT-qPCR showing the inclusion of *EYA3* exon 7 and control exons 2 and 9 in presence of DMSO or ABA. Experiments were performed with a specific RNA guide near EYA3 exon 7 or with an unrelated guide in the *ZNF618* gene (Ctrl). Data are represented as the mean value \pm S.E.M. of independent experiments ($n = 3$). t -test (* P -val < 0.05).

the DDX5 ortholog Dbp2 involved the formation of secondary structures within the 3'UTR of targeted transcripts (66). This rather points to a role of the helicase in facilitating the processing of nascent transcripts, allowing termination. In line with this, we found that a high proportion of transcripts presenting a 3' extension also present splicing alterations, suggesting a global perturbation of RNA processing of these transcripts, consistent with the recent demonstration that splicing defects are often associated with inefficient 3' end cleavage (20).

CTCF binds frequently to chromatin at or near DDX5/DDX17-dependent internal and terminal exons, and it co-regulates at least a subset of those exons. CTCF binding is promoted by DDX5/DDX17 and deleting a CTCF binding site as far as 1.2 kb downstream of a PAS is sufficient to increase transcriptional read-through of the gene, which highlights the prominent role played by CTCF. In these conditions, DDX5/DDX17 are still required for 3' end processing and termination, indicating a complex relationship between these helicases and CTCF, as discussed also below. A function of CTCF in regulating splicing and polyadenylation has been documented earlier (45–48,94), but how it does so is still unclear, as various mechanisms have been proposed (reviewed in (49)). This includes a possible effect on the nascent transcript through the RNA binding activity of CTCF (95,96), but most evidence points to an effect of CTCF on the progression of the transcription complex along the gene. CTCF is thought to create a roadblock inducing RNAPII stalling, especially when it binds downstream of alternatively spliced exons (45,50,97,98). In agreement with those results, CTCF binding sites around DDX5/DDX17-regulated exons are more often found at the level or the exon or within a 2 kb window downstream, and RNAPII density is higher across CTCF-associated exons compared to other exons.

Our results indicate that DDX5/DDX17-dependent exons, in particular those associated to CTCF binding, are often engaged in chromatin loops, which establishes a link between gene looping, alternative splicing and 3' end processing, in line with earlier reports (51,53,54). However, we now provide the first direct demonstration that gene looping is instrumental for splicing and 3' end processing regulation at the RNA level.

One remarkable finding is the high frequency of contacts between the first and the last exons of genes whose 3' end processing and/or splicing is regulated by DDX5/DDX17, which suggests that this 3D conformation is a feature of DDX5/DDX17-dependent genes. Note that the analysed ChIA-PET datasets and our experiments were performed in different biological contexts, so the link between the specific 3D organization of genes and their regulation by DDX5/DDX17 may be underestimated. The CTCF sites flanking those gene loops form an equal proportion of convergent and tandem pairs, which differs from boundaries of topologically associated domains (TADs) or long-range promoter-enhancer interactions, that largely involve convergent sites (99–103). However, these strong and stable loops represent only a fraction of CTCF-involving chromatin loops in cells, and tandem loops, which are mostly found within larger convergent loops (typically TADs), represent a third of CTCF-mediated loops (104). Their weaker

contact frequency suggests they are more dynamic and associated with transcription or regulatory functions, which fits well with our results.

Gene looping is at the same time a consequence of the transcriptional process and a facilitator of transcriptional/co-transcriptional steps (2,40), but so far most studies on gene looping have been made in yeast, and in mammalian cells the picture may differ to some extent. Gene looping is modulated according to the transcriptional status of genes, and it involves factors that are necessary for transcription activation (27–36,96,105,106). Importantly, gene looping is also an essential component of the cross-talk between transcription and co-transcriptional RNA processing, as it connects the 3' end processing machinery and the promoter region, which may ensure the proper recognition of the PAS and transcription termination (28,34,41,50). Our results confirm this idea, since we showed on two different genes a correlation between a DDX5/DDX17-dependent gene loop and the deregulation of transcription termination and 3' end processing in absence of both RNA helicases. Furthermore, we demonstrated that forcing DNA looping at a given PAS increased its use to the detriment of downstream sites. Our data suggest also that promoter-terminator gene looping is indirectly associated with splicing regulation. The inclusion of a subset of weak promoter-proximal exons, that are overrepresented within DDX5/DDX17-repressed exons, is mediated by an altered transcriptional activity near the promoter. At least in the prototypical example of *PRMT2* exon 2, this may result from the opening of the looped conformation of the gene.

We do not know yet how RNA helicases control the formation of gene loops at the molecular level, but as DDX5/DDX17 depletion reduced CTCF binding and gene looping, CTCF may be involved in these intragenic contacts. Yet, we cannot exclude the contribution of other factors, as DDX5/DDX17 were shown previously to be required for stabilizing chromatin loops in a CTCF-independent context (71). Interestingly, a recent report described the existence of several types of loops joining the 5' and 3' ends of genes regulated by estrogen or retinoic acid receptor, including loops that are stabilized by R-loops forming at both sides of genes (34). As DDX5 was reported to modulate R-loop formation at promoter or terminator regions (64,65,107), it will be important to test whether DDX5/DDX17-dependent gene looping involves RNA/DNA hybrids.

In line with this hypothesis, DDX5/DDX17-dependent exons associated to CTCF binding and high RNAPII density display a higher GC content than other exons (Figure 5). This is true for both internal and terminal exons, indicating that the mechanism could be similar in splicing and in 3' end processing. A high GC content may favour the formation of competing secondary structures in the nascent transcript, which likely explains the requirement of RNA helicases for including those exons, as shown previously (58,61,62,66). Besides, the high GC content may affect the cross-talk between transcription and RNA processing. The link between the GC content and transcription is complex since a high GC content has been associated to less RNAPII pauses *in vitro* (93), but also to correlate nega-

tively with RNAPII speed (91,92). Yet RNA structures and RNA-DNA hybrids, that are thermodynamically favoured by a high GC content, clearly have an impact on transcription *in vivo*, and *vice versa* (24,108), consistent with an active role of RNA helicases in this cross-talk. In this context, the fact that DDX5/DDX17 depletion tends to increase further the RNAPII density at CTCF-associated exons, while it decreases exon inclusion and 3' end cleavage, contradicts the 'window of opportunity' model of co-transcriptional regulation of RNA processing (7,21). However, it highlights a key function of these factors in transcriptional and co-transcriptional processes across GC-rich regions. The main function of DDX5 and DDX17 may indeed be to disentangle the RNA and/or DNA structures forming during transcription, with consequences on RNAPII dynamics and gene looping, although gene looping itself may be a contributing factor to exon inclusion. What distinguishes some genes from others in their dependence to DDX5/DDX17 is likely to reside in their propensity to form local structures that could obstruct the processing signals along the gene and the nascent RNA molecule. How DDX5/DDX17-regulated structures connect to CTCF binding and gene looping will have to be clarified, but a link between formation of R-loops and CTCF binding was recently discovered (109), and this will deserve to be explored further in light of our results.

In conclusion, by providing the first direct evidence that chromatin looping can impact alternative splicing and polyadenylation, our study changes the way we envision the regulation of RNA processing. The use of a subset of GC-rich internal and terminal exons is dependent on RNA helicases DDX5 and DDX17 and on CTCF binding to chromatin, and it is associated with DDX5/DDX17-dependent gene looping. The spatial organization of genes and its modulation will have to be taken into account to understand the roles played by DDX5 and DDX17 in the control of cell proliferation and differentiation, or the consequences of their deregulation in cancer.

DATA AVAILABILITY

The raw data for RNA-seq and RNAPII ChIP-seq have been deposited in NCBI's Gene Expression Omnibus (110) and are accessible through GEO Series accession number GSE183205 and GSE183517.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We thank Dr Kevin C. Wang (Stanford University School of Medicine) and Dr Philippe Mangeot (CIRI, Lyon) for the generous gift of plasmids. We are grateful to LBMC members for fruitful discussions and advice, especially Dr Vincent Vanoosthuysen for his critical reading of the manuscript. We acknowledge the contribution of AniRA lentivectors production facility from the CELPHEDIA Infrastructure and SFR Biosciences (UMS3444/CNRS,

US8/Inserm, ENS de Lyon, UCBL), the GenomEast sequencing platform of IGBMC (Illkirch, France) for high-throughput sequencing, and the PSMN (Pôle Scientifique de Modélisation Numérique) of the ENS-Lyon for computing resources.

FUNDING

Agence Nationale pour la Recherche [ANR-16-CE12-0009-01]; Fondation ARC [PGA120140200853]; Ligue contre le Cancer (Equipe labellisée, Comités du Rhône et de la Loire); Institut National du Cancer; AFM-Téléthon and Association Hubert Gouin 'Enfance et Cancer'; S.T., J.V., A.L. and L.B.A. received doctoral fellowships respectively from AFM-Téléthon (to S.T. and A.L.); French Ministry of Research (to J.V.); Ligue contre le Cancer (to J.V. and L.B.A.); G.G. and E.C. were supported by the Agence Nationale pour la Recherche; J.B.C. was supported by Fondation de France. Funding for open access charge: CNRS.

Conflict of interest statement. None declared.

REFERENCES

- Neugebauer, K.M. (2019) Nascent RNA and the coordination of splicing with transcription. *Cold Spring Harb. Perspect. Biol.*, **11**, a032227.
- Al-Husini, N., Medler, S. and Ansari, A. (2020) Crosstalk of promoter and terminator during RNA polymerase II transcription cycle. *Biochim. Biophys. Acta Gene Regul. Mech.*, **1863**, 194657.
- Herzel, L., Ottoz, D.S.M., Alpert, T. and Neugebauer, K.M. (2017) Splicing and transcription touch base: co-transcriptional spliceosome assembly and function. *Nat. Rev. Mol. Cell Biol.*, **18**, 637–650.
- Tellier, M., Maudlin, I. and Murphy, S. (2020) Transcription and splicing: a two-way street. *Wiley Interdiscipl. Rev. RNA*, **11**, e1593.
- Aslanzadeh, V., Huang, Y., Sanguinetti, G. and Beggs, J.D. (2018) Transcription rate strongly affects splicing fidelity and cotranscriptionality in budding yeast. *Genome Res.*, **28**, 203–213.
- Braberg, H., Jin, H., Moehle, E.A., Chan, Y.A., Wang, S., Shales, M., Benschop, J.J., Morris, J.H., Qiu, C., Hu, F. *et al.* (2013) From structure to systems: high-resolution, quantitative genetic analysis of RNA polymerase II. *Cell*, **154**, 775–788.
- de la Mata, M., Alonso, C.R., Kadener, S., Fededa, J.P., Blaustein, M., Pelisch, F., Cramer, P., Bentley, D. and Kornblihtt, A.R. (2003) A slow RNA polymerase II affects alternative splicing *in vivo*. *Mol. Cell*, **12**, 525–532.
- Fong, N., Kim, H., Zhou, Y., Ji, X., Qiu, J., Saldi, T., Diener, K., Jones, K., Fu, X.D. and Bentley, D.L. (2014) Pre-mRNA splicing is facilitated by an optimal RNA polymerase II elongation rate. *Genes Dev.*, **28**, 2663–2676.
- Ip, J.Y., Schmidt, D., Pan, Q., Ramani, A.K., Fraser, A.G., Odom, D.T. and Blencowe, B.J. (2011) Global impact of RNA polymerase II elongation inhibition on alternative splicing regulation. *Genome Res.*, **21**, 390–401.
- Maslon, M.M., Braunschweig, U., Aitken, S., Mann, A.R., Kilanowski, F., Hunter, C.J., Blencowe, B.J., Kornblihtt, A.R., Adams, I.R. and Caceres, J.F. (2019) A slow transcription rate causes embryonic lethality and perturbs kinetic coupling of neuronal genes. *EMBO J.*, **38**, e101244.
- Oesterreich, F.C., Herzel, L., Straube, K., Hujer, K., Howard, J. and Neugebauer, K.M. (2016) Splicing of nascent RNA coincides with intron exit from RNA polymerase II. *Cell*, **165**, 372–381.
- Furger, A., O'Sullivan, J.M., Binnie, A., Lee, B.A. and Proudfoot, N.J. (2002) Promoter proximal splice sites enhance transcription. *Genes Dev.*, **16**, 2792–2799.
- Alexander, R.D., Innocente, S.A., Barrass, J.D. and Beggs, J.D. (2010) Splicing-dependent RNA polymerase pausing in yeast. *Mol. Cell*, **40**, 582–593.
- Caizzi, L., Monteiro-Martins, S., Schwalb, B., Lysakovskaia, K., Schmitzova, J., Sawicka, A., Chen, Y., Lidschreiber, M. and Cramer, P.

- (2021) Efficient RNA polymerase II pause release requires U2 snRNP function. *Mol. Cell*, **81**, 1920–1934.
15. Fiszbein, A., Krick, K.S., Begg, B.E. and Burge, C.B. (2019) Exon-mediated activation of transcription starts. *Cell*, **179**, 1551–1565.
 16. Fong, Y.W. and Zhou, Q. (2001) Stimulatory effect of splicing factors on transcriptional elongation. *Nature*, **414**, 929–933.
 17. Chathoth, K.T., Barrass, J.D., Webb, S. and Beggs, J.D. (2014) A splicing-dependent transcriptional checkpoint associated with prespliceosome formation. *Mol. Cell*, **53**, 779–790.
 18. Mayer, A., di Iulio, J., Maleri, S., Eser, U., Vierstra, J., Reynolds, A., Sandstrom, R., Stamatoyannopoulos, J.A. and Churchman, L.S. (2015) Native elongating transcript sequencing reveals human transcriptional activity at nucleotide resolution. *Cell*, **161**, 541–554.
 19. Nojima, T., Gomes, T., Grosso, A.R.F., Kimura, H., Dye, M.J., Dhir, S., Carmo-Fonseca, M. and Proudfoot, N.J. (2015) Mammalian NET-Seq reveals Genome-wide nascent transcription coupled to RNA processing. *Cell*, **161**, 526–540.
 20. Reimer, K.A., Mimoso, C.A., Adelman, K. and Neugebauer, K.M. (2021) Co-transcriptional splicing regulates 3' end cleavage during mammalian erythropoiesis. *Mol. Cell*, **81**, 998–1012.
 21. Eperon, L.P., Graham, I.R., Griffiths, A.D. and Eperon, I.C. (1988) Effects of RNA secondary structure on alternative splicing of pre-mRNA: is folding limited to a region behind the transcribing RNA polymerase? *Cell*, **54**, 393–401.
 22. Dujardin, G., Lafaille, C., de la Mata, M., Marasco, L.E., Munoz, M.J., Le Jossic-Corcoss, C., Corcos, L. and Kornblihtt, A.R. (2014) How slow RNA polymerase II elongation favors alternative exon skipping. *Mol. Cell*, **54**, 683–690.
 23. Giono, L.E. and Kornblihtt, A.R. (2020) Linking transcription, RNA polymerase II elongation and alternative splicing. *Biochem. J.*, **477**, 3091–3104.
 24. Saldi, T., Riemondy, K., Erickson, B. and Bentley, D.L. (2021) Alternative RNA structures formed during transcription depend on elongation rate and modify RNA processing. *Mol. Cell*, **81**, 1789–1801.
 25. Proudfoot, N.J. (2016) Transcriptional termination in mammals: stopping the RNA polymerase II juggernaut. *Science*, **352**, aad9926.
 26. Mapendano, C.K., Lykke-Andersen, S., Kjems, J., Bertrand, E. and Jensen, T.H. (2010) Crosstalk between mRNA 3' end processing and transcription initiation. *Mol. Cell*, **40**, 410–422.
 27. 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.
 28. Mukundan, B. and Ansari, A. (2013) Srb5/Med18-mediated termination of transcription is dependent on gene looping. *J. Biol. Chem.*, **288**, 11384–11394.
 29. Singh, B.N. and Hampsey, M. (2007) A transcription-independent role for TFIIB in gene looping. *Mol. Cell*, **27**, 806–816.
 30. Medler, S., Al Husini, N., Raghunayakula, S., Mukundan, B., Aldea, A. and Ansari, A. (2011) Evidence for a complex of transcription factor IIB with poly(A) polymerase and cleavage factor I subunits required for gene looping. *J. Biol. Chem.*, **286**, 33709–33718.
 31. O'Sullivan, J.M., Tan-Wong, S.M., Morillon, A., Lee, B., Coles, J., Mellor, J. and Proudfoot, N.J. (2004) Gene loops juxtapose promoters and terminators in yeast. *Nat. Genet.*, **36**, 1014–1018.
 32. Larkin, J.D., Cook, P.R. and Papanonis, A. (2012) Dynamic reconfiguration of long human genes during one transcription cycle. *Mol. Cell Biol.*, **32**, 2738–2747.
 33. Le May, N., Fradin, D., Iltis, I., Bougneres, P. and Egly, J.M. (2012) XPG and XPF endonucleases trigger chromatin looping and DNA demethylation for accurate expression of activated genes. *Mol. Cell*, **47**, 622–632.
 34. Pezone, A., Zuchegna, C., Tramontano, A., Romano, A., Russo, G., de Rosa, M., Vinciguerra, M., Porcellini, A., Gottesman, M.E. and Avvedimento, E.V. (2019) RNA stabilizes transcription-dependent chromatin loops induced by nuclear hormones. *Sci. Rep.*, **9**, 3925.
 35. Tan-Wong, S.M., French, J.D., Proudfoot, N.J. and Brown, M.A. (2008) Dynamic interactions between the promoter and terminator regions of the mammalian BRCA1 gene. *Proc. Natl Acad. Sci. U.S.A.*, **105**, 5160–5165.
 36. Zuchegna, C., Aceto, F., Bertoni, A., Romano, A., Perillo, B., Laccetti, P., Gottesman, M.E., Avvedimento, E.V. and Porcellini, A. (2014) Mechanism of retinoic acid-induced transcription: histone code, DNA oxidation and formation of chromatin loops. *Nucleic Acids Res.*, **42**, 11040–11055.
 37. Tan-Wong, S.M., Zaugg, J.B., Camblong, J., Xu, Z., Zhang, D.W., Mischo, H.E., Ansari, A.Z., Luscombe, N.M., Steinmetz, L.M. and Proudfoot, N.J. (2012) Gene loops enhance transcriptional directionality. *Science*, **338**, 671–675.
 38. Laine, J.P., Singh, B.N., Krishnamurthy, S. and Hampsey, M. (2009) A physiological role for gene loops in yeast. *Genes Dev.*, **23**, 2604–2609.
 39. Tan-Wong, S.M., Wijayatilake, H.D. and Proudfoot, N.J. (2009) Gene loops function to maintain transcriptional memory through interaction with the nuclear pore complex. *Genes Dev.*, **23**, 2610–2624.
 40. Grzechnik, P., Tan-Wong, S.M. and Proudfoot, N.J. (2014) Terminate and make a loop: regulation of transcriptional directionality. *Trends Biochem. Sci.*, **39**, 319–327.
 41. Lamas-Maceiras, M., Singh, B.N., Hampsey, M. and Freire-Picos, M.A. (2016) Promoter-Terminator gene loops affect alternative 3'-End processing in yeast. *J. Biol. Chem.*, **291**, 8960–8968.
 42. Xiang, J.F. and Corces, V.G. (2021) Regulation of 3D chromatin organization by CTCF. *Curr. Opin. Gen. Dev.*, **67**, 33–40.
 43. Braccioli, L. and de Wit, E. (2019) CTCF: a Swiss-army knife for genome organization and transcription regulation. *Essays Biochem.*, **63**, 157–165.
 44. van Ruiten, M.S. and Rowland, B.D. (2021) On the choreography of genome folding: a grand pas de deux of cohesin and CTCF. *Curr. Opin. Cell Biol.*, **70**, 84–90.
 45. Shukla, S., Kavak, E., Gregory, M., Imashimizu, M., Shutinoski, B., Kashlev, M., Oberdoerffer, P., Sandberg, R. and Oberdoerffer, S. (2011) CTCF-promoted RNA polymerase II pausing links DNA methylation to splicing. *Nature*, **479**, 74–79.
 46. Alharbi, A.B., Schmitz, U., Marshall, A.D., Vanichkina, D., Nagarajah, R., Vellozzi, M., Wong, J.J., Bailey, C.G. and Rasko, J.E. (2021) Ctf haploinsufficiency mediates intron retention in a tissue-specific manner. *RNA Biol.*, **18**, 93–103.
 47. Lopez Soto, E.J. and Lipscombe, D. (2020) Cell-specific exon methylation and CTCF binding in neurons regulate calcium ion channel splicing and function. *Elife*, **9**, e54879.
 48. Marina, R.J., Sturgill, D., Bailly, M.A., Thenoz, M., Varma, G., Prigge, M.F., Nanan, K.K., Shukla, S., Haque, N. and Oberdoerffer, S. (2016) TET-catalyzed oxidation of intragenic 5-methylcytosine regulates CTCF-dependent alternative splicing. *EMBO J.*, **35**, 335–355.
 49. Alharbi, A.B., Schmitz, U., Bailey, C.G. and Rasko, J.E.J. (2021) CTCF as a regulator of alternative splicing: new tricks for an old player. *Nucleic Acids Res.*, **49**, 7825–7838.
 50. Nanavaty, V., Abrash, E.W., Hong, C., Park, S., Fink, E.E., Li, Z., Sweet, T.J., Bhasin, J.M., Singuri, S., Lee, B.H. *et al.* (2020) DNA methylation regulates alternative polyadenylation via CTCF and the cohesin complex. *Mol. Cell*, **78**, 752–764.
 51. Mercer, T.R., Edwards, S.L., Clark, M.B., Neph, S.J., Wang, H., Stergachis, A.B., John, S., Sandstrom, R., Li, G., Sandhu, K.S. *et al.* (2013) DNase I-hypersensitive exons colocalize with promoters and distal regulatory elements. *Nat. Genet.*, **45**, 852–859.
 52. Curado, J., Iannone, C., Tilgner, H., Valcarcel, J. and Guigo, R. (2015) Promoter-like epigenetic signatures in exons displaying cell type-specific splicing. *Genome Biol.*, **16**, 236.
 53. Ruiz-Velasco, M., Kumar, M., Lai, M.C., Bhat, P., Solis-Pinson, A.B., Reyes, A., Kleinsorg, S., Noh, K.M., Gibson, T.J. and Zaugg, J.B. (2017) CTCF-Mediated chromatin loops between promoter and gene body regulate alternative splicing across individuals. *Cell Syst.*, **5**, 628–637.
 54. Grubert, F., Srivas, R., Spacek, D.V., Kasowski, M., Ruiz-Velasco, M., Sinnott-Armstrong, N., Greenside, P., Narasimha, A., Liu, Q., Geller, B. *et al.* (2020) Landscape of cohesin-mediated chromatin loops in the human genome. *Nature*, **583**, 737–743.
 55. Bourgeois, C.F., Mortreux, F. and Auboeuf, D. (2016) The multiple functions of RNA helicases as drivers and regulators of gene expression. *Nat. Rev. Mol. Cell Biol.*, **17**, 426–438.
 56. Xing, Z., Ma, W.K. and Tran, E.J. (2019) The DDX5/Dbp2 subfamily of DEAD-box RNA helicases. *Wiley Interdiscipl. Rev. RNA*, **10**, e1519.

57. Ameur, L.B., Marie, P., Thenoz, M., Giraud, G., Combe, E., Claude, J.B., Lemaire, S., Fontrodona, N., Polveche, H., Bastien, M. *et al.* (2020) Intragenic recruitment of NF-kappaB drives splicing modifications upon activation by the oncogene tax of HTLV-1. *Nat. Commun.*, **11**, 3045.
58. Camats, M., Guil, S., Kokolo, M. and Bach-Elias, M. (2008) P68 RNA helicase (DDX5) alters activity of cis- and trans-acting factors of the alternative splicing of H-Ras. *PLoS One*, **3**, e2926.
59. Dardenne, E., Pierredon, S., Driouch, K., Grataudou, L., Lacroix-Triki, M., Espinoza, M.P., Zonta, E., Germann, S., Mortada, H., Villemain, J.P. *et al.* (2012) Splicing switch of an epigenetic regulator by RNA helicases promotes tumor-cell invasiveness. *Nat. Struct. Mol. Biol.*, **19**, 1139–1146.
60. Dardenne, E., Polay Espinoza, M., Fattet, L., Germann, S., Lambert, M.P., Neil, H., Zonta, E., Mortada, H., Grataudou, L., Deygas, M. *et al.* (2014) RNA helicases DDX5 and DDX17 dynamically orchestrate transcription, miRNA, and splicing programs in cell differentiation. *Cell Rep.*, **7**, 1900–1913.
61. Kar, A., Fushimi, K., Zhou, X., Ray, P., Shi, C., Chen, X., Liu, Z., Chen, S. and Wu, J.Y. (2011) RNA helicase p68 (DDX5) regulates tau exon 10 splicing by modulating a stem-loop structure at the 5' splice site. *Mol. Cell Biol.*, **31**, 1812–1821.
62. Lambert, M.P., Terrone, S., Giraud, G., Benoit-Pilven, C., Cluet, D., Combaret, V., Mortreux, F., Auboeuf, D. and Bourgeois, C.F. (2018) The RNA helicase DDX17 controls the transcriptional activity of REST and the expression of proneural microRNAs in neuronal differentiation. *Nucleic Acids Res.*, **46**, 7686–7700.
63. Lee, Y.J., Wang, Q. and Rio, D.C. (2018) Coordinate regulation of alternative pre-mRNA splicing events by the human RNA chaperone proteins hnRNP1 and DDX5. *Genes Dev.*, **32**, 1060–1074.
64. Katahira, J., Ishikawa, H., Tsujimura, K., Kurono, S. and Hieda, M. (2019) Human THO coordinates transcription termination and subsequent transcript release from the HSP70 locus. *Genes Cells*, **24**, 272–283.
65. Mersaoui, S.Y., Yu, Z., Coulombe, Y., Karam, M., Busatto, F.F., Masson, J.Y. and Richard, S. (2019) Arginine methylation of the DDX5 helicase RGG/RG motif by PRMT5 regulates resolution of RNA:DNA hybrids. *EMBO J.*, **35**, e100986.
66. Lai, Y.H., Choudhary, K., Cloutier, S.C., Xing, Z., Aviran, S. and Tran, E.J. (2019) Genome-Wide discovery of DEAD-Box RNA helicase targets reveals RNA structural remodeling in transcription termination. *Genetics*, **212**, 153–174.
67. Giraud, G., Terrone, S. and Bourgeois, C.F. (2018) Functions of DEAD box RNA helicases DDX5 and DDX17 in chromatin organization and transcriptional regulation. *BMB Rep.*, **51**, 613–622.
68. Fujita, T. and Fujii, H. (2011) Direct identification of insulator components by insertional chromatin immunoprecipitation. *PLoS One*, **6**, e26109.
69. Yao, H., Brick, K., Evrard, Y., Xiao, T., Camerini-Otero, R.D. and Felsenfeld, G. (2010) Mediation of CTCF transcriptional insulation by DEAD-box RNA-binding protein p68 and steroid receptor RNA activator SRA. *Genes Dev.*, **24**, 2543–2555.
70. Fuller-Pace, F.V. (2013) The DEAD box proteins DDX5 (p68) and DDX17 (p72): multi-tasking transcriptional regulators. *Biochem. Biophys. Acta*, **1829**, 756–763.
71. Morgan, S.L., Mariano, N.C., Bermudez, A., Arruda, N.L., Wu, F., Luo, Y., Shankar, G., Jia, L., Chen, H., Hu, J.F. *et al.* (2017) Manipulation of nuclear architecture through CRISPR-mediated chromosomal looping. *Nat. Commun.*, **8**, 15993.
72. Anders, S., Pyl, P.T. and Huber, W. (2015) HTSeq—a python framework to work with high-throughput sequencing data. *Bioinformatics*, **31**, 166–169.
73. Benoit-Pilven, C., Marchet, C., Chautard, E., Lima, L., Lambert, M.P., Sacomoto, G., Rey, A., Cologne, A., Terrone, S., Dulaurier, L. *et al.* (2018) Complementarity of assembly-first and mapping-first approaches for alternative splicing annotation and differential analysis from RNAseq data. *Sci. Rep.*, **8**, 4307.
74. Shen, S., Park, J.W., Lu, Z.X., Lin, L., Henry, M.D., Wu, Y.N., Zhou, Q. and Xing, Y. (2014) rMATS: robust and flexible detection of differential alternative splicing from replicate RNA-Seq data. *Proc. Natl Acad. Sci. U.S.A.*, **111**, E5593–E5601.
75. Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R. and Genome Project Data Processing, S. (2009) The sequence alignment/map format and SAMtools. *Bioinformatics*, **25**, 2078–2079.
76. Hu, B., Petela, N., Kurze, A., Chan, K.L., Chapard, C. and Nasmith, K. (2015) Biological chromodynamics: a general method for measuring protein occupancy across the genome by calibrating chip-seq. *Nucleic Acids Res.*, **43**, e132.
77. Di Tommaso, P., Chatzou, M., Floden, E.W., Barja, P.P., Palumbo, E. and Notredame, C. (2017) Nextflow enables reproducible computational workflows. *Nat. Biotech.*, **35**, 316–319.
78. Mallinoud, P., Villemain, J.P., Mortada, H., Polay Espinoza, M., Desmet, F.O., Samaan, S., Chautard, E., Tranchevent, L.C. and Auboeuf, D. (2014) Endothelial, epithelial, and fibroblast cells exhibit specific splicing programs independently of their tissue of origin. *Genome Res.*, **24**, 511–521.
79. Lemaire, S., Fontrodona, N., Aube, F., Claude, J.B., Polveche, H., Modolo, L., Bourgeois, C.F., Mortreux, F. and Auboeuf, D. (2019) Characterizing the interplay between gene nucleotide composition bias and splicing. *Genome Biol.*, **20**, 259.
80. van Heeringen, S.J. and Veenstra, G.J. (2011) GimmeMotifs: a de novo motif prediction pipeline for chip-sequencing experiments. *Bioinformatics*, **27**, 270–271.
81. Castro-Mondragon, J.A., Riudavets-Puig, R., Rauluseviciute, I., Lemma, R.B., Turchi, L., Blanc-Mathieu, R., Lucas, J., Boddie, P., Khan, A., Manosalva Perez, N. *et al.* (2022) JASPAR 2022: the 9th release of the open-access database of transcription factor binding profiles. *Nucleic Acids Res.*, **50**, D165–D173.
82. Quinlan, A.R. and Hall, I.M. (2010) BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*, **26**, 841–842.
83. Bailey, T.L. and Elkan, C. (1994) Fitting a mixture model by expectation maximization to discover motifs in biopolymers. *Proc. Int. Conf. Intell. Syst. Mol. Biol.*, **2**, 28–36.
84. Barski, A., Cuddapah, S., Cui, K., Roh, T.Y., Schones, D.E., Wang, Z., Wei, G., Chepelev, I. and Zhao, K. (2007) High-resolution profiling of histone methylations in the human genome. *Cell*, **129**, 823–837.
85. Hyle, J., Zhang, Y., Wright, S., Xu, B., Shao, Y., Easton, J., Tian, L., Feng, R., Xu, P. and Li, C. (2019) Acute depletion of CTCF directly affects MYC regulation through loss of enhancer-promoter looping. *Nucleic Acids Res.*, **47**, 6699–6713.
86. Khoury, A., Achinger-Kawecka, J., Bert, S.A., Smith, G.C., French, H.J., Luu, P.L., Peters, T.J., Roh, Q., Parry, A.J., Valdes-Mora, F. *et al.* (2020) Constitutively bound CTCF sites maintain 3D chromatin architecture and long-range epigenetically regulated domains. *Nat. Commun.*, **11**, 54.
87. Kubo, N., Ishii, H., Xiong, X., Bianco, S., Meitinger, F., Hu, R., Hocker, J.D., Conte, M., Gorkin, D., Yu, M. *et al.* (2021) Promoter-proximal CTCF binding promotes distal enhancer-dependent gene activation. *Nat. Struct. Mol. Biol.*, **28**, 152–161.
88. Luan, J., Xiang, G., Gomez-Garcia, P.A., Tome, J.M., Zhang, Z., Vermunt, M.W., Zhang, H., Huang, A., Keller, C.A., Giardine, B.M. *et al.* (2021) Distinct properties and functions of CTCF revealed by a rapidly inducible degron system. *Cell Rep.*, **34**, 108783.
89. Nora, E.P., Goloborodko, A., Valton, A.L., Gibcus, J.H., Uebersohn, A., Abdennur, N., Dekker, J., Mirny, L.A. and Bruneau, B.G. (2017) Targeted degradation of CTCF decouples local insulation of chromosome domains from genomic compartmentalization. *Cell*, **169**, 930–944.
90. Zuin, J., Dixon, J.R., van der Reijden, M.I., Ye, Z., Kolovos, P., Brouwer, R.W., van de Corput, M.P., van de Werken, H.J., Knoch, T.A., van, I.W.F. *et al.* (2014) Cohesin and CTCF differentially affect chromatin architecture and gene expression in human cells. *Proc. Natl. Acad. Sci. U.S.A.*, **111**, 996–1001.
91. Jonkers, I., Kwak, H. and Lis, J.T. (2014) Genome-wide dynamics of pol II elongation and its interplay with promoter proximal pausing, chromatin, and exons. *Elife*, **3**, e02407.
92. Veloso, A., Kirkconnell, K.S., Magnuson, B., Biewen, B., Paulsen, M.T., Wilson, T.E. and Ljungman, M. (2014) Rate of elongation by RNA polymerase II is associated with specific gene features and epigenetic modifications. *Genome Res.*, **24**, 896–905.
93. Zamft, B., Bintu, L., Ishibashi, T. and Bustamante, C. (2012) Nascent RNA structure modulates the transcriptional dynamics of RNA polymerases. *Proc. Natl. Acad. Sci. U.S.A.*, **109**, 8948–8953.

94. Monahan, K., Rudnick, N.D., Kehayova, P.D., Pauli, F., Newberry, K.M., Myers, R.M. and Maniatis, T. (2012) Role of CCCTC binding factor (CTCF) and cohesin in the generation of single-cell diversity of protocadherin-alpha gene expression. *Proc. Natl. Acad. Sci. U.S.A.*, **109**, 9125–9130.
95. Hansen, A.S., Hsieh, T.S., Cattoglio, C., Pustova, I., Saldana-Meyer, R., Reinberg, D., Darzacq, X. and Tjian, R. (2019) Distinct classes of chromatin loops revealed by deletion of an RNA-Binding region in CTCF. *Mol. Cell*, **76**, 395–411.
96. Saldana-Meyer, R., Rodriguez-Hernandez, J., Escobar, T., Nishana, M., Jacome-Lopez, K., Nora, E.P., Bruneau, B.G., Tsiganos, A., Furlan-Magaril, M., Skok, J. *et al.* (2019) RNA interactions are essential for CTCF-Mediated genome organization. *Mol. Cell*, **76**, 412–422.
97. Agirre, E., Bellora, N., Allo, M., Pages, A., Bertucci, P., Kornblihtt, A.R. and Eyra, E. (2015) A chromatin code for alternative splicing involving a putative association between CTCF and HP1alpha proteins. *BMC Biol.*, **13**, 31.
98. Paredes, S.H., Melgar, M.F. and Sethupathy, P. (2013) Promoter-proximal CCCTC-factor binding is associated with an increase in the transcriptional pausing index. *Bioinformatics*, **29**, 1485–1487.
99. de Wit, E., Vos, E.S., Holwerda, S.J., Valdes-Quezada, C., Verstegen, M.J., Teunissen, H., Splinter, E., Wijchers, P.J., Krijger, P.H. and de Laat, W. (2015) CTCF binding polarity determines chromatin looping. *Mol. Cell*, **60**, 676–684.
100. Vietri Rudan, M., Barrington, C., Henderson, S., Ernst, C., Odom, D.T., Tanay, A. and Hadjur, S. (2015) Comparative Hi-C reveals that CTCF underlies evolution of chromosomal domain architecture. *Cell Rep.*, **10**, 1297–1309.
101. Sanborn, A.L., Rao, S.S., Huang, S.C., Durand, N.C., Huntley, M.H., Jewett, A.I., Bochkov, I.D., Chinnappan, D., Cutkosky, A., Li, J. *et al.* (2015) Chromatin extrusion explains key features of loop and domain formation in wild-type and engineered genomes. *Proc. Natl. Acad. Sci. U.S.A.*, **112**, E6456–E6465.
102. Guo, Y., Xu, Q., Canzio, D., Shou, J., Li, J., Gorkin, D.U., Jung, I., Wu, H., Zhai, Y., Tang, Y. *et al.* (2015) CRISPR inversion of CTCF sites alters genome topology and enhancer/promoter function. *Cell*, **162**, 900–910.
103. Rao, S.S., Huntley, M.H., Durand, N.C., Stamenova, E.K., Bochkov, I.D., Robinson, J.T., Sanborn, A.L., Machol, I., Omer, A.D., Lander, E.S. *et al.* (2014) A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell*, **159**, 1665–1680.
104. Tang, Z., Luo, O.J., Li, X., Zheng, M., Zhu, J.J., Szalaj, P., Trzaskoma, P., Magalska, A., Wlodarczyk, J., Ruszczycki, B. *et al.* (2015) CTCF-Mediated human 3D genome architecture reveals chromatin topology for transcription. *Cell*, **163**, 1611–1627.
105. Hsieh, T.S., Cattoglio, C., Slobodyanyuk, E., Hansen, A.S., Rando, O.J., Tjian, R. and Darzacq, X. (2020) Resolving the 3D landscape of transcription-linked mammalian chromatin folding. *Mol. Cell*, **78**, 539–553.
106. Rowley, M.J., Lyu, X., Rana, V., Ando-Kuri, M., Karns, R., Bosco, G. and Corces, V.G. (2019) Condensin II counteracts cohesin and RNA polymerase II in the establishment of 3D chromatin organization. *Cell Rep.*, **26**, 2890–2903.
107. Villarreal, O.D., Mersaoui, S.Y., Yu, Z., Masson, J.Y. and Richard, S. (2020) Genome-wide R-loop analysis defines unique roles for DDX5, XRN2, and PRMT5 in DNA/RNA hybrid resolution. *Life Sci. All.*, **3**, e202000762.
108. Turowski, T.W., Petfalski, E., Goddard, B.D., French, S.L., Helwak, A. and Tollervy, D. (2020) Nascent transcript folding plays a major role in determining RNA polymerase elongation rates. *Mol. Cell*, **79**, 488–503.
109. Luo, H., Zhu, G., Eshelman, M.A., Fung, T.K., Lai, Q., Wang, F., Zeisig, B.B., Lesperance, J., Ma, X., Chen, S. *et al.* (2022) HOTTIP-dependent R-loop formation regulates CTCF boundary activity and TAD integrity in leukemia. *Mol. Cell*, **82**, 833–851.
110. Edgar, R., Domrachev, M. and Lash, A.E. (2002) Gene expression omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res.*, **30**, 207–210.