Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

RNA pol II speed controls the level of read-through mediated transcriptional interference

Virginie Jouffret*
University of Toulouse, CNRS, Centre de Biologie Intégrative (CBI), Toulouse, France
https://orcid.org/0000-0001-9435-4974

Maurane Eisele*
University of Toulouse, CNRS, Centre de Biologie Intégrative (CBI), Toulouse, France

Tomas Caetano
University of Toulouse, CNRS, Centre de Biologie Intégrative (CBI), Toulouse, France

Marika Iacovone
University of Toulouse, CNRS, Centre de Biologie Intégrative (CBI), Toulouse, France

Julie Bossuyt
University of Toulouse, CNRS, Centre de Biologie Intégrative (CBI), Toulouse, France

Coralie Delpech
University of Toulouse, CNRS, Centre de Biologie Intégrative (CBI), Toulouse, France

Marion Aguirrebengoa
University of Toulouse, CNRS, Centre de Biologie Intégrative (CBI), Toulouse, France
https://orcid.org/0000-0001-5205-9774

Maxime Delmas
University of Toulouse, CNRS, Centre de Biologie Intégrative (CBI), Toulouse, France

Estelle Nicolas
University of Toulouse, CNRS, Centre de Biologie Intégrative (CBI), Toulouse, France
https://orcid.org/0000-0003-0412-8477

Didier Trouche**
University of Toulouse, CNRS, Centre de Biologie Intégrative (CBI), Toulouse, France

Lisa Muniz**
lisa.muniz@utoulouse.fr

University of Toulouse, CNRS, Centre de Biologie Intégrative (CBI), Toulouse, France
https://orcid.org/0000-0001-5809-2444

Article

Keywords:

Posted Date: November 3rd, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10623247/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: There is **NO** Competing Interest.

Abstract

Transcriptional interference, which can be mediated by read-through transcription, is recognized as a mechanism negatively controlling gene expression levels. Recent studies highlighted the importance of transcriptional interference mediated by read-through transcription in response to different kinds of stresses and in response to viral infections. In this study we investigated the extent of transcriptional interference mediated by read-through transcription in different human cell lines. Our results, based on the analysis of transcriptomics and genomics data, reveal that transcriptional interference is widespread in the human genome in unperturbed conditions. Furthermore, through the use of RNA pol II mutants with different velocity, we show that transcriptional interference can be controlled by RNA pol II speed. Finally, we observe that read-through transcription invading downstream genes significantly associates with decreased level of translation of downstream genes in tandem gene pairs, therefore probably affecting protein levels. Together these data indicate that transcriptional interference is a mechanism controlling gene expression in unperturbed conditions. In addition, our results uncover the importance of transcription speed in controlling the expression level of downstream genes in tandem gene pairs.

***Virginie Jouffret and Maurane Eisele contributed equally to this work.**

****Didier Trouche and Lisa Muniz contributed equally to this work.**

Introduction

In eukaryotes transcription by RNA polymerase II controls the production of messenger-RNA (mRNA) as well as several non-coding RNA. Transcription by RNA pol II is tightly regulated at all stages of its cycle i.e., initiation, promoter proximal pausing, elongation and termination. Transcription termination is defined as the dissociation of the ternary complex formed by the RNA pol II, the nascent RNA and the DNA matrix. On protein coding genes, transcription termination is tightly linked to processing of the 3' end of mRNAs. It begins with the transcription of the polyA signal that induces conformational changes in the RNA pol II elongation complex, most probably linked to dephosphorylation events. These conformational changes then lead to a slow-down of the RNA pol II which allow its catch-up by the 5' to 3' exonuclease Xrn2 (1, 2). In higher eukaryotes, transcription termination by RNA pol II is not taking place at a precise location downstream the gene but in a more diffuse manner in a so called 'termination window' (3). Transcriptional read-through refers to the continuation of transcription beyond this termination window.

Transcription termination is a dynamically regulated process which can respond to different stimuli. Its control is especially important as it can affect not only the expression of the gene that is transcribed itself but also the expression of the gene located downstream on the genome, whether it is a tandem gene or a convergent gene. Indeed, if the RNA pol II does not terminate transcription before reaching the gene located downstream, it can cause transcriptional interference, which is defined as 'the suppressive

influence of one transcription process, directly and in cis on a second transcriptional process' (4). Transcriptional interference can occur through different mechanisms including promoter competition, "sitting duck" interference, occlusion, collision and roadblock (reviewed in (4)).

Transcriptional interference has been shown to occur in many different organisms. For example, in *S. cerevisiae* transcription of the SER3 gene, involved in the first step of serine and glycine biosynthesis, is tightly repressed during growth in rich medium. This repression has been shown to involve a transcriptional interference-based mechanism in which transcription of the SRG1 non coding RNA across the SER3 promoter is associated with a high nucleosome occupancy and interferes with the binding of activators to the promoter of the SER3 gene (5, 6). In *S. pombe*, three phosphate acquisition genes are repressed in phosphate rich conditions by transcriptional interference through transcription of long non coding RNA genes located upstream of these three different protein-coding genes (7–9). In this case, transcriptional elongation factors travelling with the polymerase that initiated transcription on the upstream lncRNA promoter establish a repressive chromatin environment over the downstream protein-coding gene promoters preventing the binding of transcription factors (8). In *D. melanogaster*, the expression of bxd ncRNAs represses the expression of homeobox gene Ubx by transcriptional interference and participates in the establishment of the mosaic pattern of Ubx during embryonic development (10). In mouse embryonic stem cells, transcription of the Airn lncRNA over the Igf2r promoter silences the expression of the Igf2r imprinted gene by reducing the recruitment of functional RNA pol II to its promoter (11).

In addition, transcriptional interference has been shown to be coupled to translational repression of the mRNA produced from the downstream promoter, in pairs of alternative promoters in yeast as well as in human cells (12, 13). Indeed, transcription from the upstream promoter leads to the production of an mRNA harboring a 5' leader sequence containing an upstream ORF that affects the translation efficiency of the downstream ORF. This phenomenon depends on Swi/Snf co-transcriptional nucleosome remodeling around the downstream promoter (14).

While all the above examples of transcriptional interference refer to long non coding RNA genes regulating the expression of downstream protein-coding gene promoters, or to alternative promoters of a same gene, several studies have shown that transcriptional interference can also be mediated by read-through transcription from protein-coding genes. A recent study in ant neurons has shown that activation of a single odorant receptor gene in a large tandem array of many related genes silences downstream genes expression through transcriptional interference mediated by read-through transcription which is crucial to ensure specificity of the olfactory sensory neurons (15, 16). Along the same line, it has been shown in human cells that deletion of the promoter of the CSDE1 protein coding gene induced an increase in the expression of the N-Ras gene located downstream (17), probably also reflecting transcriptional interference mediated by transcriptional read-through from the CSDE1 gene. Finally, we have shown a few years ago that activation of transcriptional read-through on some genes in cells undergoing oncogene induced senescence is associated with a decrease in the expression of convergent genes (18). We showed for two of these read-through RNAs, that their increased expression

is mediated by an increase in the RNA pol II speed in these intergenic regions, and that they repress the expression of the downstream gene.

Besides these examples showing read-through mediated transcriptional interference on one or a few protein coding genes in physiological conditions, recent studies highlighted the possible extent of this mechanism in response to different kinds of stresses, in response to viral infections as well as in cancer (19–25). Indeed, in these studies, transcriptional read-through was shown to invade the genes located downstream (called read-in), to produce fusion transcripts and alter the chromatin state and 3D organization, which all could potentially affect the transcription of the genes located downstream (19–25). Notably, it has been shown that the translation efficiency of the invaded gene mRNAs can be affected (20, 22).

Interestingly, transcriptional read-through has recently been shown to be widespread in the human genome (26), which raises the question of the consequences of such events on downstream genes' expression. Indeed, although discovered more than forty-five years ago (27–29) and the subject of numerous studies, transcriptional interference and its potential as an important mechanism of gene expression regulation remain elusive. Most of the studies showing transcriptional interference have been focusing on a single or a few specific gene pairs. As a consequence, the extent of transcriptional interference in eukaryotic genomes remains unknown.

Here, we investigated the extent of transcriptional interference in human cells, its link with read-through transcription and how it could affect downstream genes translation.

Results

Read-through transcription is widespread in the human genome

When analyzing the different classes of genes in RNAseq data from WI38 cells, we noticed that while both genes in pairs of convergent or divergent genes were expressed at the same level, within a pair of tandem genes the downstream gene tended to be less expressed than the upstream gene (18). This striking observation led us to analyse this bias in more details and in different cell lines. We first analysed the expression of tandem genes, that we defined as genes organized in the same direction and separated by a maximal distance of 100kb, in another non-cancerous cell line (MCF10A ER-Src (30) hereafter called MCF10A) as well as in cancer cells (U2OS) (31). From total RNA-seq data, we plotted metagene profiles and boxplots for all tandem genes and found that the difference between expression of the upstream and downstream genes also holds true in these two cell lines (Fig. 1A and supplementary Fig. 1). We thought that this could be due to transcriptional interference mediated by read-through transcription coming from the upstream gene. To test this hypothesis, we used an in-house developed pipeline to detect read-through transcripts from RNAseq data. Briefly, this pipeline, that we named HMM, uses a Hidden Markov Model to detect continuous signal after the annotated polyA site

from bigwig files (see materials and methods). Before using this pipeline, we tested its performance by comparing the read-through transcripts detection with two other published pipelines DOGfinder (32) and ARTdeco (21). We detected read-through transcripts from our four replicates of published U2OS total RNAseq data using default parameters for ARTdeco and DOGfinder. ARTdeco, DOGfinder and HMM detected 2783, 6439 and 5617 read-through transcripts expressed in U2OS cells respectively. ARTdeco detected read-through transcripts produced from 13.8% of all genes while DOGfinder and HMM detected read-through transcripts from about 25% of all genes. We found a good overlap between read-through transcripts detected by the three pipelines (supplementary Fig. 2A). The HMM pipeline detected most of the read-through transcripts detected by both published pipelines: 72% of the read-through transcripts detected by ARTdeco and almost 51% of the read-through transcripts detected by DOGfinder. Moreover, 25.5% of all the read-through transcripts were uniquely detected by the HMM pipeline. When we compared the expression level of genes producing read-through transcripts (+ RT) or not (-RT) identified by the three different pipelines, we noticed that the HMM pipeline actually detects read-through transcripts produced from genes that are less expressed as compared to read-through transcripts identified by the DOGfinder and ARTdeco pipelines (supplementary Fig. 2B), suggesting that it is less biased by the expression levels of genes from which read-through transcripts originate. In conclusion, we found that HMM has a similar sensitivity than ARTdeco and DOGfinder and is slightly better in detecting read-through transcripts from lowly expressed genes. We therefore used the HMM pipeline in the rest of the study.

We detected read-through transcripts originating from expressed genes (FPKM > 1) from two RNAseq replicates in U2OS, MCF10A and WI38 cell lines. We identified 2560, 3238, and 2844 read-through transcripts as compared to 12300, 11025, and 9354 expressed genes in U2OS, WI38 and MCF10A respectively, indicating that between 21% and 30% of expressed genes produce read-through transcripts in these cell lines. Interestingly, we found that 52% of read-through transcripts are found in at least two of the cell lines and 21% are common to all three cell lines (supplementary Fig. 2C), suggesting that some genes are more prone to produce read-through transcripts than others, as recently shown in human tissues (26).

Read-through transcription limits the expression of downstream genes in tandem and convergent gene pairs

In order to test whether transcriptional interference mediated by read-through transcription could be responsible for the lower expression of downstream genes in pairs of tandem genes, we separated pairs of tandem genes into two different classes using total RNAseq data from U2OS cells: one for which the upstream gene produces read-through (+ RT) and the other not (-RT). We then drew metagene profiles and boxplots for these two classes of tandem gene pairs, as well as boxplots representing the ratio of expression of downstream genes versus upstream genes (Fig. 1B and supplementary Fig. 3A). While the expression level of the downstream gene is about the same as the expression level of the upstream gene in the -RT class, the expression of the downstream gene is lower than the expression of the

upstream gene in the + RT class. We also drew the same metagene profile and boxplots but using the read-through detected by either DOGfinder or ARTdeco. We found the same trend towards a lower expression of the downstream genes in + RT gene pairs (supplementary Fig. 3B), indicating that read-through transcription is associated with a lower expression of downstream genes whatever the algorithm used to detect them. However, we also noted that we tend to enrich the + RT genes class with genes that are expressed at a higher level as compared to -RT genes, even when using the HMM algorithm although it can detect read-through from lowly expressed genes as compared to DOGfinder and ARTdeco (see supplementary Fig. 2B). This observation can be explained by the fact that the detection of read-through transcripts necessitates a minimal level of coverage on the read-through region and thus a minimal level of expression of the gene producing the read-through. But, it could lead to a bias when comparing the expression level of downstream genes with the expression level of upstream genes in pairs of tandem genes in the -RT and + RT classes. In order to avoid this bias, we decided to directly compare the level of expression of downstream genes from the -RT and + RT classes (Fig. 1C), which showed that the level of expression of downstream genes in the + RT class is lower than the level of expression of downstream genes in the -RT class.

To test whether this finding holds true in other cell lines, we drew metagene profiles (supplementary Fig. 4A) as well as boxplots to directly compare the level of expression of downstream genes from the -RT and + RT classes (supplementary Fig. 4B) using total RNAseq data from MCF10A and WI38 cells. The results obtained from MCF10A and WI38 cells recapitulated the results obtained from U2OS cells. Taken together, these results suggest that read-through transcription from the upstream gene interferes with expression of the downstream gene in the + RT class.

Finally, we thought that read-in transcription might bias total RNAseq data by generating reads on the downstream genes in tandem gene pairs. Indeed, read-in transcription has been shown to induce an apparent increase in expression of downstream genes in conditions in which read-through transcription is highly activated such as IAV and HSV-1 infection (19–21) as well as different stress conditions (22). Transcriptional interference might thus be underestimated in total RNAseq data. PolyA+ RNAseq data are probably less biased given that read-in transcripts might not always be properly matured or exported to the cytoplasm and will be less represented. To test this hypothesis, we analysed published polyA+ data from U2OS cells (33, 34) (Fig. 1D) and found that the difference in expression of downstream genes in the + RT class as compared to expression of downstream genes in the -RT class is more pronounced in polyA+ RNAseq data than in total RNAseq data (Fig. 1C), thus indicating that mature transcripts may be strongly affected by transcriptional interference.

We wondered whether such an interference mediated by read-through transcription could also be detected on convergent gene pairs. When we previously drew metagene profiles with our RNAseq data from WI38 cells, we noted that the expression levels of convergent genes were lower than that of tandem or divergent genes and postulated that it could reflect a higher tendency of transcriptional interference between convergent genes (18). However, in this study, we did not take into account whether genes produced read-through or not. We thus sorted convergent gene pairs in -RT and + RT

classes and then divided the + RT class into two classes according to the strand on which the gene producing read-through is located. We then drew metagene profiles for the different classes of genes in U2OS cells (Fig. 1E). These analyses indeed revealed that when a gene from a convergent gene pair produces read-through, the convergent gene is less expressed, consistent with transcriptional interference as for tandem gene pairs. Again, we performed the same analyses on RNAseq data from WI38 and MCF10A cells and found the same trend towards a lower expression of the convergent gene when the other gene in the pair produces read-through (supplementary Fig. 5A). As for tandem genes, we noted that we tend to enrich the + RT gene class with genes that are expressed at a higher level as compared to -RT genes, which could lead to a bias when comparing genes producing read-through with genes that do not produce read-through. In order to avoid this bias, we decided to compare the level of expression of genes convergent to genes producing read-through to that of genes convergent to genes not producing read-through. Genes located on the plus and minus strands were analysed separately in order to take into account the strand on which the gene producing read-through is located (supplementary Fig. 5B). These analyses showed that genes convergent to genes producing read-through tend to be less expressed than genes convergent to genes that do not produce read-through, again in agreement with the interpretation that transcriptional read-through interferes with the expression of downstream convergent genes.

(A) Metagene analysis of strand specific total RNA-seq profiles representing the median within the gene population of the normalized coverage in tandem genes and intergenic regions in U2OS and MCF10A cells. Analyses are shown for one replicate out of two for MCF10A cells and out of four for U2OS cells. Amongst these loci only pairs with a mean of the normalized coverage computed on the gene body > 0 and an intergenic region size of more than 100 nt and less than 100kb were considered (4707 loci for U2OS cells and 4071 loci for MCF10A cells). **(B)** Same as in (A) for U2OS cells, except that tandem gene pairs were separated into two classes based on the production or not of read-through transcripts (RT) from the upstream gene detected by the HMM algorithm, as indicated. Analysis of one representative RNA-seq replicate is shown out of four. A minimum threshold on upstream genes expression was applied to -RT gene class in order to avoid biasing our analysis due to a difference in the level of expression of upstream genes in between the -RT and + RT classes (3704 gene pairs in the -RT class; 987 gene pairs in the + RT class). **(C)** Boxplots showing the median, first and third quantiles of the median normalized coverage for downstream genes from the -RT and + RT classes in tandem gene pairs from total RNAseq data from U2OS cells. The -RT and + RT tandem genes classes were selected as in (B). Analyses are shown for one replicate out of four. *$p < 0.05$ by Wilcoxon-Mann-Whitney test and adjusted by the Benjamini-Hochberg procedure. **(D)** Tandem gene pairs were separated into two classes based on the production of read-through transcripts from the upstream gene (-RT and + RT) in published polyA+ RNAseq data from U2OS cells (33, 34) and the level of expression of downstream genes was determined for each class. The -RT and + RT tandem genes classes were selected as in (B). Boxplots show the median, first and third quantiles of the level of expression of these genes expressed in CPM. ***$p < 0.01$ by Wilcoxon-Mann-Whitney test and adjusted by the Benjamini-Hochberg procedure. **(E)** Same as in (B) except that convergent genes' loci were analysed and separated based on the strand on

which the gene producing read-through transcripts is located (1044, 356, 575, and 555 loci in the -RT, +RT on both strands, +RT on the minus strand, and + RT on the plus strand classes, respectively).

Read-through transcription limits transcription of downstream genes in tandem gene pairs

Since we performed these analyses using total and polyA+ RNAseq data, we cannot exclude the possibility that the effect we detect on the downstream gene in tandem and convergent gene pairs is post-transcriptional. To show that the interference is at the transcription level, we tested RNA pol II loading on downstream genes when the upstream gene produces read-through. We thus repeated the same kind of analyses but using published CTD Serine 5 phosphorylated (Ser5P) RNA pol II ChIPseq data. Indeed, CTD phosphorylated on Serine 5 marks promoter region and is necessary for transcription initiation. To perform this analysis, we used two published datasets obtained in the non-cancerous HEK293 cells and cancer cells DLD1 (35, 36). We first identified read-through transcripts from either RNAseq or tNETseq data obtained from the same datasets in order to classify tandem genes in the -RT and + RT classes. As in the other cell lines we analysed above, the expression of the downstream gene is lower in the + RT class as compared to the -RT class (supplementary Fig. 6A), suggesting that read-through transcription from the upstream gene also interferes with expression of the downstream gene. Strikingly, the RNA pol II CTD Ser5P profiles from both cell lines show that read-through transcription from the upstream gene correlates with a decrease in initiating RNA pol II loading on the promoter of the downstream gene (supplementary Fig. 6B). However, these metagene profiles also show that the level of RNA pol II CTD Ser5P is higher on + RT genes than on -RT genes, suggesting that these genes are indeed more transcribed. We thus also compared directly the level of RNA pol II CTD Ser5P loaded on the promoter of downstream genes in the -RT and + RT genes classes (Fig. 2A). These results show that genes located downstream of genes producing read-through have less initiating RNA pol II loaded on their promoters as compared to genes located downstream of genes that do not produce read-through. These data suggest that the reduction in RNA levels from the downstream gene within the + RT gene class is transcriptional rather than post-transcriptional.

In order to confirm that the effect we detect on downstream gene's expression level is transcriptional, we also tested whether it holds true when analysing nascent RNAs. We thus performed six replicates of TTchemseq experiments in U2OS cells. We then used the HMM pipeline to detect read-through transcripts from these nascent RNAseq data and analysed the transcription level of downstream genes in the -RT and + RT classes (Fig. 2B). Although, as expected due to read-in transcription, not significant when analysing each experiment independently, analysis of all six replicates together revealed that transcription of downstream genes in the + RT class is lower than transcription of downstream genes in the -RT class. Taken together these results show that the lower expression of downstream genes within the + RT gene class is associated with a lower transcription of these genes and is most probably due to transcriptional interference mediated by read-through transcription.

(A) Analysis of the median normalized coverage of S5P RNA pol II ChIPseq data from DLD1 (36) and HEK293 cells (35). Tandem gene pairs were separated into two classes based on the production of read-through transcripts (RT) from the upstream gene or not as indicated. Amongst these loci only pairs with a mean of the normalized coverage computed on the gene body > 0 and an intergenic region size of more than 100nt and less than 100kb were considered. The level of RNA pol II Ser5P detected on downstream genes was determined for each class. Boxplots show the median, first and third quantiles of the median normalized coverage of the ChIPseq on these genes. **(B)** Tandem gene pairs were separated into two classes based on the production of read-through transcripts from the upstream gene or not (-RT and + RT) in TTchemseq data from U2OS cells and the level of expression of downstream genes was determined for each class. Left panel: Analyses are shown for one replicate out of six. Boxplots show the median, first and third quantiles of the median normalized coverage for these genes. ns: non-significant. Right panel: The dot plots show the medians of the median normalized coverage for downstream genes in the -RT and + RT classes from six independent experiments, with individual experiments identified by circles of different colors. *p < 0.05 by paired Welch test.

The level of transcriptional interference depends on the distance travelled by RNA pol II downstream the polyA site

If transcriptional interference is mediated by polymerases coming from the upstream gene entering into the downstream gene promoter, then the level of transcriptional interference would depend on the distance travelled by RNA pol II after the polyA site of the upstream gene. To test this hypothesis, we designed several reporter systems in which we introduced two genes organized in tandem and in which the transcription termination signal of the upstream gene is identical but the distance separating the two genes varies. Indeed, we hypothesized that varying the distance between the two genes but keeping the same transcription termination efficiency would mimic a change in the distance travelled by RNA pol II after the polyA site. We thus cloned two reporter genes (mCherry and GFP) organized in tandem under the control of identical strong promoters (CMV) (Fig. 3A). We used the termination signal from the human BYSL gene for transcription termination of the upstream gene mCherry. Indeed, in U2OS cells, transcription termination from the BYSL gene takes place in a zone of around 1kb to 2kb downstream of the polyA site (data not shown). We constructed five different reporter systems in which the two reporter genes are separated by 500pb (0.5kb-IntR), 1kb (1kb-IntR), 2kb (2kb-IntR), 3kb (3kb-IntR) or 4kb (4kb-IntR) of the intergenic region located immediately downstream of the BYSL gene. We transfected U2OS cells with these different constructs and analysed by flow-cytometry the level of expression of the mCherry and GFP proteins. Strikingly, we found that GFP protein expression increased with the distance between the mCherry and GFP genes (Fig. 3B, left panel and supplementary Fig. 7B), a behaviour also observed compared to mCherry as illustrated by the GFP/Cherry ratio (supplementary Fig. 7A and B).

At the RNA level, GFP expression does not change between the 0.5kb-IntR reporter and the 2kb-IntR reporter, while it increases when the two genes are separated by 3kb or more (Fig. 3B, right panel). Although surprising considering GFP protein expression results, this finding could be explained by the fact that when measuring GFP RNA levels in the 0.5kb-IntR, 1kb-IntR and 2kb-IntR we actually detect

read-in transcription, i.e., transcription coming from the mCherry gene and entering into the GFP gene, which decreases when the distance between the tandem genes increases. To test this possibility, we analysed by RT-qPCR a region of the CMV promoter located upstream from the GFP gene. We found that transcription in this region is high in the 0.5kb-IntR and decreases to become nearly undetectable in the 4kb-IntR, anti-correlating with the distance separating the tandem genes (Fig. 3C). These results suggest that the GFP RNAs amplified by RT-qPCR are a mixture of RNAs started at the TSS of the GFP and fusion RNAs containing both the mCherry sequence and the GFP sequence. However, since there is no IRES to promote translation of the GFP ORF from the fusion RNA, the GFP protein cannot be produced from these fusion RNAs, which leads to a decrease in its expression level.

Taken together, these results thus show that in a setting in which the transcription termination signal is identical, the distance separating the genes determines the expression level of the downstream gene in the pair of tandem genes, as expected for transcriptional interference mediated by read-through transcription.

We next wondered whether in a setting in which the transcription termination signal varies from gene to gene such as the human genome, the distance separating the genes could still affect the level of transcriptional interference. We thus plotted the level of expression of the downstream gene against the distance separating it from the upstream gene for the + RT gene pairs (Fig. 3D). These results show that in the genome the level of expression of downstream genes does not only depend on the distance separating the genes.

(A) Schematic representation of the reporter system used to analyse transcriptional interference. The mCherry and GFP coding sequences were cloned downstream of two identical CMV promoters and separated by genomic sequences of different lengths beginning at a position located 109bp upstream of the polyA site of the human BYSL gene. **(B)** Left panel: GFP protein expression level measured by flow cytometry in U2OS cells transfected with the different reporter systems. The dot plots show the medians for 10 000 cells from three independent experiments, with individual experiments identified by circles of different colors. ns: non-significant, *p < 0.05, ***p < 0.01 by paired Welch test. Right panel: GFP RNA expression level measured by RT-qPCR in U2OS cells transfected with the different reporter systems, normalized by the level of expression of the puromycine resistance gene which is expressed from the reporter plasmid in order to account for transfection efficiency differences in between samples. The histogram shows the means and standard deviation from three independent experiments. *p < 0.05, ns: non-significant by Student's T-test. **(C)** Expression level of CMV promoter RNAs measured by RT-qPCR in U2OS cells transfected with the different reporter systems normalized by the level of expression of puromycin resistance gene expressed from the reporter plasmids. Means and standard deviation from three independent experiments are shown. *p < 0.05, ***p < 0.01, ns: non-significant by Student's T-test. **(D)** Dot plot representing the level of expression (FPKM) of downstream genes from the + RT tandem genes' class according to the intergenic distance separating them from the upstream gene expressed in bp.

RNA pol II speed can control the level of interference on the downstream gene in tandem gene pairs

Transcription speed has been shown to control the distance travelled by RNA pol II after the polyA site of a gene, with slower transcription shifting termination closer to the polyA site while faster transcription extends the distance travelled by RNA pol II after the polyA site (37). We thus hypothesized that slowing-down RNA pol II may decrease the transcriptional interference on downstream genes in a context in which it is high while accelerating RNA pol II may increase the transcriptional interference on downstream genes in a context in which it is low. We thus co-transfected the different reporter systems (0.5kb-IntR, 1kb-IntR, 2kb-IntR, 3kb-IntR or 4kb-IntR) together with plasmids encoding alpha-amanitin resistant slow (R749H), fast (E1126G) or WT RNA pol II (38) and measured the level of GFP and Cherry expression by flow cytometry (see Fig. 4A for GFP expression and supplementary Fig. 8 for GFP/mCherry ratio). In agreement with our hypothesis, expression of the slow mutant induces an increase in the GFP expression from reporter systems on which transcriptional interference is high (0.5kb-IntR, 1kb-IntR, 2kb-IntR), suggesting that transcriptional interference is reduced. In contrast, the fast mutant expression has no or little effect on the level of GFP expression on these reporter systems, probably because transcriptional interference is already maximal.

Strikingly, expression of the fast RNA pol II mutant induces a decrease in the GFP expression from reporter systems on which transcriptional interference is low (3kb-IntR and 4kb-IntR), suggesting that transcriptional interference is increased (Fig. 4A and supplementary Fig. 8). In contrast, the slow mutant has little or no effect on the level of GFP expression on these reporter systems, probably because transcriptional interference is already minimal. Thus, taken together these results indicate that transcriptional interference is determined by RNA pol II speed.

Since these experiments using mutant RNA pol II have been performed on transiently transfected reporter systems, we wondered whether these results would hold true in a chromatin context. So, we stably integrated the 1kb-IntR and the 4kb-IntR reporter systems in the AAVS1 safe harbor site of U2OS cells, thus establishing two independent cell lines containing either the 1kb-IntR or the 4kb-IntR. We noticed that as for transiently transfected reporter systems, the level of expression of RNAs at the CMV promoter is lower in the 4kb-IntR cell line as compared to the 1kb-IntR cell line, while the level of expression of the GFP is much higher (supplementary Fig. 9).

We then analysed by flow cytometry the level of expression of GFP and mCherry in these two cell lines after transient transfection of the RNA pol II speed mutants. As for the transiently transfected reporter system, expression of the slow mutant induces an increase in the GFP expression level in the 1kb-IntR cell line but has little effect in the 4kb-IntR cell line (Figs. 4B, 4C and supplementary Figs. 10 and 11). The fast RNA pol II mutant on the other hand, has little effect on the expression of GFP in the 1kb-IntR cell line (Fig. 4B, left panel, supplementary Fig. 10), whereas it induces a decrease in this expression in the 4kb-IntR cell line (Fig. 4C, left panel and supplementary Fig. 11). The effect of the RNA pol II speed mutants on GFP expression level is recapitulated at the RNA level (Figs. 4B and 4C, right panels), suggesting that the effect is transcriptional.

Thus, altogether, these data indicate that transcriptional interference on downstream genes is controlled by RNA pol II speed, probably through its effect on transcriptional read-through.

(A) GFP protein expression levels were measured by flow cytometry in U2OS cells co-transfected with plasmids encoding either the 0.5kb-IntR, 1kb-IntR, 2kb-IntR, 3kb-IntR or 4kb-IntR reporter systems together with the WT, Slow or Fast RNA pol II mutants expression vectors and treated with alpha-amanitin for 48 hours. Dot plots show the median for 10 000 cells in each condition, from three independent experiments identified by circles of different colors. *p < 0.05, ***p < 0.01, ns: non-significant by Welch t-test and adjusted by the Benjamini-Hochberg procedure. **(B)** Left panel: The 1kb-IntR reporter cell line was transfected with plasmids encoding the WT, Slow or Fast RNA pol II mutants and treated with alpha-amanitin for 48 hours. The GFP protein expression levels were measured by flow cytometry. Dot plots show the median for 10 000 cells in each condition, from four independent experiments identified by circles of different colors. ***p < 0.01, *p < 0.05 by Welch t-test. Right panel: GFP RNAs expression levels were measured by RT-qPCR and normalized relative to GAPDH mRNA expression level. The means and standard deviations from four independent experiments are shown. ***p < 0.01, ns: non-significant by Student's T-test. **(C)** Same as in (B) except that the 4kb-IntR reporter cell line was used. Left panel: Dot plots show the median for 10 000 cells in each condition, from six independent experiments identified by circles of different colors.

Read-through mediated transcriptional interference can be regulated by perturbations of the speed of endogenous RNA pol II

To directly test whether changes in the GFP RNA levels are due to transcription coming from the mCherry gene, we intended to stop RNA pol II originating from the mCherry gene by targeting a dCas9 to the non-transcribed DNA strand in the intergenic region between mCherry and GFP. Indeed, this has been shown to efficiently block the progression of bacterial RNA polymerase (39) and to suppress read-through transcription when targeted downstream a polyA site in human cells (40). To avoid steric hindrance between the dCas9 targeted in the intergenic region and the loading of RNA pol II onto the CMV promoter of GFP, the guide RNA was designed to target a site located 2kb upstream from the CMV promoter of GFP in the 4kb-IntR reporter cell line. Upon expression of the fast RNA pol II mutant to generate transcriptional interference on this reporter system, we found that targeting a dCas9 to the intergenic region induces an increase in GFP protein expression level (Fig. 5A left panel, supplementary Fig. 12). The effect of the targeting of dCas9 in the intergenic region on GFP expression level is recapitulated at the RNA level (Fig. 5A middle panel). Moreover, when measuring the CMV promoter RNAs level we found that, as expected if transcription is blocked in the intergenic region, its expression is lower after targeting the dCas9 in the intergenic region (Fig. 5A, right panel). These results thus demonstrate that the transcriptional interference we detect in the stably integrated reporter system is due to RNA pol II originating from the upstream gene and travelling through the intergenic region.

Previous experiments showing the effect of RNA pol II speed on transcriptional interference were performed using mutants of RNA pol II. We next intended to test whether changes in RNA pol II speed

observed upon environmental changes can also affect transcriptional interference.

RNA pol II speed is increased upon histone hyper-acetylation induced by the use of the histone deacetylase (HDAC) inhibitor TSA (41), therefore creating a more open chromatin structure. Another way to affect RNA pol II speed is through the inhibition of the PP1 phosphatase which serves as a brake for RNA pol II throughout the genome. Although its effect is more important in termination regions, PNUTS-PP1 inhibition induces a 1.6 fold increase in RNA pol II speed in gene bodies and a 3.4 fold increase in termination regions (42). We thus tested the effect of TSA and of a specific inhibitor of PP1, tautomycetin, on the 4kb-IntR reporter cell line since it is the reporter for which we detect an important effect of the increase in the speed of RNA pol II (see Fig. 4C). Treatment of the 4kb-IntR reporter cell line with TSA induced a major increase in expression of both mCherry and GFP (data not shown and supplementary Fig. 13) probably because a more open chromatin structure favors transcription. Nevertheless, treatment with both Tautomycetin and TSA induced a strong decrease in the ratio between GFP and mCherry expression (Fig. 5B, left panel), recapitulating the effect of the fast RNA pol II mutant as expected. The effect of TSA and Tautomycetin on GFP expression level is recapitulated at the RNA level (Figs. 5B, middle panel), suggesting that the effect is transcriptional. Finally, we detected a significant increase in the level of CMV promoter RNAs after treatment with both Tautomycetin and TSA (Figs. 5B, right panel), consistent with increased read-through transcription which could be responsible for the effects we detect on GFP expression.

Thus, taken together, these data show that manipulating the speed of RNA pol II can affect the expression of genes most probably through regulating transcriptional interference.

(A) The 4kb-IntR reporter cell line was co-transfected with plasmids encoding the Fast RNA pol II mutant, a dCas9 protein, and either a control guide RNA (Scr) or a guide RNA targeting the non-transcribed strand in the intergenic region between mCherry and GFP (IntR). Left panel: GFP protein expression level was measured by flow cytometry for each individual cell in the population for 10 000 cells in each condition in each experiment. Dot plots show the medians from 3 independent experiments identified by circles of different colors. *p < 0.05 by Student's t-test. Middle and Right panels: GFP and CMV promoter RNAs expression levels were measured by RT-qPCR and normalized relative to GAPDH mRNA expression level. The means and standard deviations from three independent experiments are shown. *p < 0.05, by Student's t-test. **(B)** Left panel: GFP and mCherry proteins expression levels were measured by flow cytometry in the 4kb-IntR cell line treated with either the histone deacetylase inhibitor trichostatin A (TSA) at 3.3 µM, an inhibitor of the PP1 phosphatase Tautomycetin at 500nM or DMSO as a control. The ratio of GFP over mCherry expression was calculated for each individual cell in the population. Dot plots show the medians for 10 000 cells from four independent experiments identified by circles of different colors. *p < 0.05, ***p < 0.01 by Welch t-test. Middle and Right panels: GFP/mcherry ratio and CMV promoter RNAs expression levels were measured by RT-qPCR and normalized relative to GAPDH mRNA expression level. The means and standard deviations from five independent experiments are shown. *p < 0.05, ***p < 0.01 by Student's t-test.

The level of transcriptional interference depends on the speed of RNA pol II on endogenous genes

We then tested whether this effect of RNA pol II speed on interference on the downstream gene would hold true for endogenous tandem genes. We thus selected two pairs of tandem genes on which we can detect read-through produced from the upstream gene using the HMM algorithm (+ RT), as well as two pairs of tandem genes for which we cannot detect any read-through (-RT). For both the + RT and -RT classes we chose tandem genes separated by either a short intergenic region of 12kb (RBM25-PSEN1, +RT) and 15kb (SERPINE1-AP1S1, -RT) or by a longer intergenic region of 84kb (YAP1-BIRC3, +RT) and 86kb (SLC9A1-TENT5B, -RT) (Fig. 6A). We transfected plasmids encoding alpha-amanitin resistant slow, fast or WT RNA pol II mutants in U2OS cells. We noticed that transcription by both the fast and slow RNA pol II mutants induced a small decrease in upstream gene expression in the tandem gene pairs we analysed. We thus monitored changes in expression of the read-through relative to the upstream gene's expression. As expected, transcription by the slow RNA pol II mutant decreased the level of expression of read-through transcript from both RBM25 and YAP1 as compared to the WT RNA pol II, while transcription by the fast RNA pol II mutant increased the expression of read-through from these genes (Fig. 6B). Interestingly, while transcription by the slow RNA pol II mutant induced a small increase in the level of expression of the downstream genes in the + RT pairs as compared to transcription by the WT RNA pol II it had little or no effect on the expression of downstream genes in the -RT pairs (Fig. 6C). Conversely, transcription by the fast RNA pol II mutant lowered the expression of the two downstream genes from the + RT pairs while it had no or little effect on the downstream genes in the -RT pairs (Fig. 6C). These results suggest that RNA pol II speed can modulate the level of transcriptional interference on downstream genes in endogenous tandem gene pairs.

(A) IGV genome browser views of total RNAseq data from U2OS cells for the RBM25-PSEN1 and YAP1-BIRC3 tandem gene pairs which are part of the + RT class and SERPINE1-AP1S1 and SLC9A1-TENT5B tandem gene pairs which are part of the -RT class. The black arrows indicate the orientation of the genes. The red dotted line indicates the read-through transcript detected by the HMM algorithm for the RBM25 and YAP1 genes. The blue double arrows indicate the position and length of the intergenic region. **(B)** U2OS cells were transfected with plasmids encoding the WT, Slow or Fast RNA pol II mutants and treated with alpha-amanitin for 48 hours. The expression level of read-through transcripts was analysed by RT-qPCR for the + RT genes; the amplicon is located 11.8kb downstream of RBM25 polyA site and 0.7kb upstream of PSEN1 TSS for the RBM25-PSEN1 pair and 9.5kb downstream of YAP1 polyA site for the YAP1-BIRC3 pair. The means and standard deviations from eight independent experiments are shown. ***p < 0.01, ns: non-significant by Student's t-test. **(C)** Same as in (B) except that the level of expression of downstream genes for each tandem gene pair was measured by RT-qPCR and normalized to GAPDH mRNA expression level. The means and standard deviations from eight independent experiments are shown. *p < 0.05, ***p < 0.01, ns: non-significant by Student's t-test.

Transcriptional interference regulates cellular processes

Read-in transcription has been shown to result in a defect in translation of the invaded genes in response to HSV-1 infection (20) as well as in response to stress (22). In order to test whether this reduced translation of genes invaded by read-through is also true in non-stressed conditions, we analysed two different published ribosome profiling datasets from U2OS cells (33, 34). We calculated the ratio of the ribosome profiling signal over the polyA+ RNAseq signal and found that, in both datasets, the translation of downstream genes from tandem gene pairs in the + RT class is significantly lower than that of downstream genes in the -RT class (Fig. 7A). These results suggest that even in unperturbed conditions, transcriptional interference lowers the translation of downstream genes.

We then wondered whether genes whose expression can be regulated by read-through transcription, ie, downstream genes in tandem gene pairs and genes that are convergent to a gene producing read-through would be enriched in any specific biological process. We performed gene ontology analysis of common downstream genes from the + RT class in tandem and convergent gene pairs in U2OS, MCF10A and WI38 cells. This analysis revealed that they are enriched in genes involved in regulation of transcription by RNA pol II (Fig. 7B), which suggests that this read-through mediated transcriptional interference mechanism could be a way to fine tune the level of transcription in cells.

(A) Published ribosome profiling data from U2OS cells were analysed and normalised to polyA+ RNAseq data from the same dataset (33, 34). Tandem gene pairs were separated into two classes based on the production of read-through transcripts from the upstream gene or not (-RT and + RT) and the ratio of ribosome profiling normalized by polyA + RNA seq level of downstream genes was determined for each class. Analyses are shown for one replicate out of three. Boxplots show the median, first and third quantiles of the ratio for these genes. ***$p < 0.01$ by Wilcoxon-Mann-Whitney test and adjusted by the Benjamini-Hochberg procedure. **(B)** Gene ontology enrichment analysis performed using ShinyGO on genes located downstream tandem and convergent genes producing read-through in common between the U2OS, WI38 and MCF10A cell lines.

Discussion

Our results show that almost one fourth of all expressed genes produce read-through in the different cell lines we analysed in this study, including cancerous and non-cancerous cell lines. These results show that read-through transcription is not only a hallmark of stress response but is more widespread than previously anticipated as recently shown by different studies (26, 43). Caldas and colleagues (26) observed that between 9 and 18% of expressed genes produced read-through in healthy human tissues while we estimate that a higher percentage of genes produce read-through in unperturbed conditions. This could be explained by the fact that their analyses were limited by available data for human tissues that are non-stranded and do not present an optimal coverage for detection of read-through, thus most probably leading to an underestimation of the number of genes producing read-through. Alternatively, *in vitro* culture conditions produce some basal stress that result in permanent expression of read-through transcripts.

The organization of genes in the human genome does not seem to be under any kind of selection pressure since about one fourth are organized in a convergent orientation, one fourth are organized in a divergent orientation, one half are organized in tandem with half of them localized on the plus strand and half of them on the minus strand. However, we noticed that read-through production seems to be more biased by the organization of genes in the genome. Around 24% of tandem genes and 35% of convergent genes produce read-through transcripts. This observation could be related to the chromatin status of the intergenic region following these different classes of genes. Indeed, upstream genes in tandem gene pairs are followed by promoter regions of the downstream genes which could restrict the progression of polymerases in these regions. On the other hand, our analyses show that a higher percentage of convergent genes produce read-through as compared to other genes organization. Nevertheless, most of the polymerases producing these read-through transcripts seem to terminate as soon as they reach the end of the downstream gene likely through head-on collision of polymerases as shown in bacteria (33). Indeed, head-on collision between bacterial RNA polymerases has recently been shown to drive transcription termination in between convergent gene pairs (44). This model was proposed to represent a conserved transcription termination mechanism for convergent genes given that this gene organization is found in different organisms (44). If this mechanism also operates in human cells, we would expect a more pronounced drop in the level of reads aligning in the read-through region when reaching the convergent gene 3’ end, thus highlighting transcription termination through head-on collisions. In a previous study, we indeed observed a drop in read coverage when a class of senescence triggered read-through transcripts reaches their convergent gene 3’ end (18). By analysing metagene profiles from highly expressed read-through transcripts in U2OS and WI38 cells, we noticed that we indeed detect a drastic drop in the number of reads aligning to the read-through region in convergent gene pairs as soon as the read-through transcript reaches the convergent gene 3’ end (for example see “+RT on both strands” metaprofiles in Fig. 1C for U2OS cells and supplementary Fig. 5A for MCF10A and WI38 cells). These results suggest that as shown for bacterial RNA polymerases, this head-on collision mechanism could drive transcription termination of RNA pol II that do not terminate before reaching the convergent gene 3’ end. This could explain why the regulation of transcription termination is less tight on these convergent genes, since the cells evolved another way to terminate read-through transcription.

While Caldas and colleagues observed widespread transcriptional read-through in the human genome (26), they found that the read-through they detected invades the downstream gene at a low frequency. Our data, on the other hand, suggest that invasion of downstream genes is not so infrequent since, using ARTdeco, we detect read-in on around 11% of expressed genes. This discrepancy could come from the fact that Caldas and colleagues used polyA+ RNAseq data which could confound this kind of analyses. Indeed, when the polymerase invades the downstream gene it might not always terminate transcription at a canonical polyA site and is usually associated with incomplete splicing (20) which results in aberrant transcripts that are most probably unstable and thus undetectable by polyA+ RNAseq. However, transcription of both tandem genes by a polymerase initiating transcription at the upstream promoter can also lead to the production of a fusion transcript that can give rise to a chimeric transcript by splicing of exons from the upstream gene with exons from the downstream gene (23, 45). Chimeric

transcripts have been shown to be associated with different kinds of cancers (23, 46) but can also be found in non-diseased tissues (47) and can in some cases be translated into chimeric proteins (46). All these arguments show that while it represents an important aspect of the stress response, read-through transcription may also account for an important part of the transcriptome composition in normal tissues. Besides the possibility of giving rise to chimeric transcripts (23, 47), transcriptional read-through has been suggested to play a role in the maintenance of the nuclear stiffness which can be important in response to osmotic stress (48). In addition, these read-through transcripts have also been suggested to play a role as miRNA sponges and could thus regulate by another mean the expression level of genes (26).

The results we present in this article show that, in addition to these functions, read-through transcription can have important consequences on gene expression. Indeed, the expression of downstream genes located on the opposite strand in the case of convergent genes or on the same strand in the case of tandem genes can be affected by the expression of the upstream gene. These data suggest that the human genome has evolved so that the level of transcription of one gene can affect the expression of neighboring genes even in unperturbed conditions. Our metagene profiles indeed show that transcriptional read-through is globally associated with a lower expression level of downstream genes.

Our results suggest that read-in transcription can lead to mis-interpretation of RNAseq data when considering downstream genes in tandem gene pairs as has been shown in the case of IAV infection and heat-shock response (21). However, we show here that this read-in transcription should also be considered as an important regulator of transcription level of downstream genes in non-stress conditions. A recent study has shown that closely spaced tandem genes (separated by less than 2kb) tend to be co-expressed although RNA pol II bearing CTD modifications associated with transcription termination accumulate at the promoter region of the downstream gene (49). These results suggest that regions separating these tandem genes and especially promoters of the downstream genes have evolved to enable transcription regardless of read-in by RNA pol II coming from the upstream gene. However, this does not seem to be the case for all tandem genes as we show here that transcription of the downstream gene, regardless of the distance separating the genes tend to be lower than that of upstream genes.

Our analyses of ribosome profiling data show that downstream genes in pairs of + RT tandem gene are less translated than downstream genes of the -RT class, thus suggesting that transcriptional interference is a way to control the expression of some genes in the genome. Intriguingly, gene ontology analysis of common downstream genes in the + RT class showed an enrichment of genes involved in RNA pol II transcription. It is thus tempting to speculate that transcriptional interference mediated by read-through transcription could be a way to fine tune transcription by RNA pol II in human cells, allowing them to adapt their response to acute transcriptional changes, such as upon stress.

More generally, it would be interesting to investigate the role of pairs of tandem and convergent genes affected by this transcriptional interference mechanism to check whether they are involved in related

pathways. Indeed, transcriptional interference could be a way to couple the expression of genes involved in a pathway. For example, the YAP1 and BIRC3 genes, which form a pair of tandem genes for which we showed that RNA pol II speed can affect the expression of the downstream BIRC3 gene (Fig. 4), are both playing a role in apoptosis regulation.

Our results also provide insights into the regulation of transcriptional interference by read-through transcription. Our tests showed that the distance travelled by RNA pol II after the gene is an important determinant of transcriptional interference. Using different reporter systems we showed that in a context in which the transcription termination signal is the same, manipulating the distance between the tandem genes can affect the expression of the downstream gene. However, our analysis in the human genome showed that the level of expression of downstream genes does not only depend on the distance separating it from the upstream gene. This is probably because, the distance travelled by RNA pol II downstream genes depends on several parameters including the efficiency of transcription termination. Indeed, we showed that transcriptional interference depends on transcription speed which has been shown to affect transcription termination efficiency (37, 50). To our knowledge, this is the first demonstration that transcription speed is a major determinant of transcriptional interference, thus regulating the expression level of genes in humans. Importantly, previous findings in fission yeast showed that a slow RNA pol II mutant induces premature transcription termination of an upstream non coding RNA which relieves transcriptional interference on the downstream tandem gene (51), suggesting that this role is conserved during evolution. This result suggests that an evolutional pressure on genome organization adapted the extent of transcriptional interference between neighboring genes.

A recent study has shown that the 2kb upstream as well as the 2kb downstream of the polyA site of genes producing read-through are associated with the H3K36me3 and H3K79me2 active chromatin marks (26). These two active chromatin marks positively correlate with RNA pol II speed (52–54) again pointing to a role of RNA pol II speed in controlling the distance travelled by RNA pol II downstream the polyA site and thus influencing transcriptional interference. We have shown that histone hyperacetylation through treatment with TSA induces an increase in transcriptional interference. Together with the fact that specific histone marks are associated with read-through transcription our data suggest that modifications of the chromatin landscape could control transcriptional interference through the regulation of RNA pol II speed, adding another layer of complexity to the control of the expression levels of specific genes by chromatin modifications.

Materials and methods

RNAseq

For RNA-seq in MCF10A cells, 5–10 µg of total RNA, extracted as described below, was submitted to EMBL-GeneCore, Heidelberg, Germany. Two independent biological replicates were sequenced. We used strand-specific RNA-seq method, relying on UTP incorporation in the second cDNA strand. RNA-seq samples were sequenced using Illumina NextSeq 2000 sequencer, paired-end, 150-bp reads.

TTchem-seq

TTchem-seq was performed as described in the protocol from Gregersen et al, 2021 (55). Briefly, 2 million U2OS cells were plated in 3 15-cm dishes. 24 hours later, 4SU was added to the culture medium to a final concentration of 1mM for 10 minutes before collecting the cells in Trizol. Total RNA extraction was then performed, total RNAs were mixed with yeast 4TU-labeled RNA to serve as spike-in and fragmented using NaOH. Fragmented RNAs were then biotinylated and 4SU-RNA were pulled-down using streptavidin beads. Strand specific libraries (NEBNext Ultra II Directional RNA Library Prep Kit for Illumina, E7760S) were then prepared for high-throughput sequencing.

Bioinformatics analyses

Total RNA-seq analyses

Total RNA-seq data obtained from published studies (GSE131579 for U2OS cells (31); GSE85085 for WI38 (18); GSE145368 for DLD1 cells (35); GSE97827 for HEK293 cells (34)) were downloaded as raw sequencing files (FASQ files) and then analysed as data we have generated. Potential contamination by remaining rRNA in the ribo-depleted total RNA were removed by an initial alignment of paired-end reads to the human rDNA sequence hg38 using STAR v2.7.10a (56) with the parameters --genomeSAindexNbases 2 --outSAMunmapped Within. Unmapped reads were extracted using samtools view v1.9 (56) with flags -F 256 -f 4 -h, then realigned to the female hg38 genome with STAR v2.7.10a with default parameters. Aligned reads were sorted and indexed using samtools, retaining only primary alignments (-F 256). Normalized coverage tracks (bigWig files) were generated using deeptools bamCoverage v3.5.1 with the following parameters -bs 1 --normalizeUsing CPM --scaleFactor 50. Quality control was performed using FastQC v0.11.5 on both raw FASTQ files and final BAM files.

Ser5P-RNA pol II ChIP-seq

RNA Pol II Ser5p ChIP-seq datasets GSE97827 (34) and GSE145368 (35) were retrieved from ChIP-Atlas (58) as bigWig files after alignment on hg38.

tNETseq

RNA pol II tNet-seq datasets were obtained from GEO for WT and slow RNA pol II mutants (GSE149018 (57)). Raw sequencing data (FASTQ files) were retrieved using fastq-dump v3.0.6. Paired-end reads were aligned on hg38 using STAR v2.5.3a with parameters "--readFilesCommand gunzip -c". Aligned reads were sorted and indexed using samtools v1.9, filtered out low quality mapping (-q 25) on first dataset. Normalized coverage tracks were generated using deeptools bamCoverage v3.4.3 with the following parameters -bs 1 --normalizeUsing CPM. Raw coverage tracks (bigWig files) were generated in the same way but using parameter --normalizeUsing None instead. Quality control was performed using FastQC v0.11.5 on both raw FASTQ files and final BAM files.

Read-through detection

In total RNA-seq data, the coverage of a read-through originating from a gene is typically lower than exon coverage but similar to intron coverage. To identify read-through transcripts from such data, we thus intended to find regions downstream of genes and with a coverage similar to intron coverage of the upstream gene. We used a Continuous-State Hidden Markov Model (CS-HMM) to infer from RNA-seq coverage hidden genic states (estimated from average intron coverage) and intergenic states with a Poisson and zero-inflated Poisson mixture model. Based on RefSeq annotation and raw bigWig files, detection uses 200-bp windows (binLength) to smooth local variations. Genic states identified downstream of genes, within 500 bp (allowedRange) and with a minimum length of 5 kb (minSize), are classified as putative read-through transcripts.

In order to test the quality of these predictions, two hypotheses are tested: 1) whether the two-state model (genic/intergenic) significantly outperforms a null single-state model (p-value cutoff = 0.001, Cut_off_pvalue), and 2) whether a single-transition model is better than a random-transition model (p-value cutoff = 0.001, Cut_off_pvalue_seq). Putative read-through transcripts that do not meet these criteria are filtered out from the final list of read-through transcripts. For each data sets, detection is performed per condition, and the union of all genomic regions identified as read-through transcripts in at least one condition is returned.

The tool was applied to MCF10A, WI38, U2OS, DLD1 and HEK293T raw bigWig generated with Deeptools bamCoverage.

Tandem and convergent gene pairs annotation

From RefSeq annotation file, gene set was cleaned by removing genes entirely contained within another gene, and the longer annotated isoform of each gene was retained. 18919 genes were analysed. When ordered by start position, pairs of adjacent genes on the same strand (plus or minus) were classified as tandem genes (4920 gene pairs on strand -, 5093 gene pairs on strand +), pairs of adjacent genes on opposite strand were classified as convergent genes when arranged end-to-end (5020 gene pairs) and as divergent genes when arranged head-to-head (4612 gene pairs). No threshold on the distance between adjacent gene pairs was used. Convergent gene pairs with less than 100 bp between their transcription end sites were excluded from analyses. R 4.2.1 was used with library: rtracklayer_1.56.1, GenomicAlignments_1.32.1, plyranges_1.16.0, tidyverse_1.3.2.

Coverage filtering and metaprofiles

Normalized bigWig files were used to analyze coverage across gene pairs. For each pair, we partitioned the following regions into 20 equal bins and computed the mean coverage per bin for both strands: 1) the region upstream the gene spanning from TSS-4 kb to TSS, 2) the first gene region from TSS to TES, 3) the intergenic region between the paired genes, 4) the second gene region from TSS to TES, and 5) the downstream region from TES to TES + 4 kb.

For tandem gene pairs, when a read-through transcript was detected on the upstream gene, the pair was annotated as + RT, in a strand-specific way. For convergent gene pairs, genes producing a read-through

transcript were individually annotated as + RT based on the strand on which they are located; if both genes produced read-through transcripts, they were classified under the '+RT on both strands' category.

Pairs were removed from the analyses if either gene exhibited a sum coverage of zero (FilterGeneCov0). For tandem genes, we also filtered out lowly expressed -RT pairs using the minimum coverage of all upstream genes from the + RT class as a threshold, in order to limit the bias for highly expressed genes introduced by read-through detection (FilterGeneSupMinRT).

Metaprofiles were generated by calculating the median coverage of each bin across all genes, using R 4.2.1 with library : rtracklayer_1.56.1, GenomicAlignments_1.32.1, plyranges_1.16.0, tidyverse_1.3.2, rstatix_0.7.1.

Benchmark of ARTdeco and DOGFinder

We detected read-through transcripts in strand-specific RNA-seq data from U2OS cells (31) using three different tools: DoGFinder (default + -s flag), ARTDeco (default), and our HMM-based pipeline described above using hg38 cleanRefSeq. Events were considered consistent across methods if the predicted read-through transcripts overlapped in the TES-TES + 2 kb region, thereby classifying the corresponding genes as + RT. For supplementary Fig. 1B, gene coverage was computed using Deeptools computeMatrix (58) with the following parameters "scale-regions --binSize 200 --regionBodyLength 4000 --afterRegionStartLength 4000 --beforeRegionStartLength 4000 --skipZeros --missingDataAsZero". Metaprofiles were subsequently generated in R using the resulting coverage matrices as described above.

Published ribosome profiling data

Raw sequencing data (FASTQ files) were retrieved from the NCBI Gene Expression Omnibus (GEO) (GSE51584 (45); GSE66929 (44)) using fasterq-dump (v3.0.0). For C.rp_rep2 that had been sequenced across two lanes, the FASTQ files were merged using samtools merge to create a single file per sample prior to downstream processing. Adapter and quality trimming was carried out with the following parameters: -e 0.1 (maximum allowed error rate of 10%), -q 20 (quality trimming threshold), and -O 1 (minimum adapter overlap length). A minimum read length threshold of 18 nucleotides (Trim Galore default) was applied, and only reads passing the trimming and quality filters were retained for downstream analysis. Filtered reads were aligned to the human reference genome hg38 using STAR (v2.5.3a).

Only uniquely mapping reads were retained, and a maximum mismatch rate of 5% was allowed.

All other parameters were kept at their default settings. Read quantification was performed using featureCounts from the Rsubread package (v2.20.0). Reads were counted in single-end, unstranded mode using the RefSeqCurated GTF annotation, restricting counting to annotated exons. Multi-mapped reads were excluded, and only reads overlapping features by at least 20 nucleotides and covering at

least 90% of the read length were counted. Raw gene-level counts were normalized using counts-per-million (CPM) as implemented in edgeR (v4.0.16).

Reporter plasmids cloning

The pMK231 plasmid (#105924, Addgene) was linearised with BglII and HpaI. The Cherry and CMV-GFP sequences were amplified by PCR from the pmCherry-C1 and pEGFP-N3 plasmids respectively using primers containing 25nt overlapping ends with the linearized pMK231 plasmid (see “Cherry” and “GFP” primers in supplementary table 1). The two PCR fragments were then inserted into the linearized pMK231 using the proligation free kit from Applied biological following the manufacturer’s instructions. The resulting pMK231-mCherry-GFP plasmid was digested with BglII and HindIII in order to linearize it in between the end of the Cherry gene and the start of the CMV promoter located upstream of GFP. The BYSL (Bystin like) gene’s polyA site together with pieces of the intergenic region of different sizes were amplified by PCR from U2OS cells’ genomic DNA (see “BYSL out” primers in supplementary table 1). A nested PCR was performed on the fragment obtained with primers containing 25nt overlaps with the 3’ end of the mCherry sequence and the 5’ end of the CMV promoter located upstream of GFP (see “BYSL-xxbp” primers in supplementary table 1). Ligation of the BYSL intergenic sequence into the linearized pMK231-mCherry-GFP plasmid was performed using the HiFi DNA Assembly mastermix from NEB. For creation of the stable cell lines, the puromycin resistance gene was then replaced by the blasticidin resistance gene from the pMK391 (#121192, Addgene). The puromycine resistance gene was removed from the pMK231 plasmid by PCR using primers designed to have around 25nt overlapping ends with the blasticidin resistance gene that was extracted by PCR from the pMK391 plasmid. The two pieces were assembled using the proligation free cloning kit following the manufacturer’s instructions.

Guide RNA expressing plasmid cloning

The sequence targeting the intergenic region in between the mCherry and GFP genes in the reporter system (see supplementary table 1 for the sequence of the sgRNA BYSL) was cloned into the pESgRNA plasmid according to the protocol established in the Lachaud’s laboratory (59).

Stable integration of the reporter system in U2OS cells

The reporter systems were stably integrated into the genome of U2OS cells by homologous recombination targeted to the AAVS1 safe harbor site with a guide RNA and Cas9 protein expressed from the plasmid #72833 (Addgene). 150 000 U2OS cells were transfected with 300ng of #72833 plasmid and 700ng of the donor plasmid containing either the 1kb-IntR or the 4kb-IntR reporter systems. Cells were then treated with puromycin and diluted to obtain colonies originating from a single cell. The colonies were then tested for integration of the reporter system in the AAVS1 safe harbor site by PCR with primers targeting both the 5’ and 3’ sites of integration, but also the plasmid backbone in order to eliminate clones in which integration of the reporter system happened elsewhere in the genome.

Cell culture

MCF10A ER-Src cells were maintained as described previously for MCF10A cells (60). U2OS cells were grown in DMEM high glucose Glutamax supplemented with Sodium pyruvate, Penicilin streptomycin and 10% fetal bovine serum at 37°C and 5% CO2. Plasmids transfection was performed using the JetPrime reagent (Ozyme) according to the manufacturer's recommendations, the total amount of DNA transfected per wells of a six wells plate was 1µg using 2µL of JetPrime. The medium was changed 4 hours following transfection to limit cell toxicity. Cells were harvested 72h following transfection and either used directly for flow cytometry or for RNA extraction.

RNA pol II mutants RPB1-WT, RPB1-E1126G (Fast) and RPB1-R749H (Slow) encoding plasmids (38) were transfected into cells 24h before treatment with alpha-amanitine at a final concentration of 2.5µg/µL, cells were then collected 48h following addition of alpha-amanitine.

Cells were harvested and directly used for flow cytometry 24h following addition of TSA or tautomycetin.

For blocking the progression of RNA pol II in the intergenic region, the 4kb-IntR stable cell line was co-transfected with a plasmid encoding the fast RNA pol II mutant (RPB1-E1126G) and a plasmid encoding the dCas9 (#61355 Addgene) together with a plasmid encoding either the scramble guideRNA or the BYSL guideRNA.

RNA extraction and RT-qPCR

RNA extraction was performed using the MasterPure RNA extraction kit following the manufacturer's instructions except that the DNAse treatment was performed for 45 minutes using a mixture of RiboZero DNase and DNAse I. Reverse transcription was perfomed on 500ng of total RNA using the Superscript III RT and random primers and following the manufacturer's instructions. The qPCR was perfomed using Takara premix exTaq Sybr green mastermix and Biorad CFX instruments. See supplementary table 1 for the list and sequences of the primers used for qPCR analyses.

Flow cytometry

Cells were analyzed on a Cytoflex flow cytometer (Beckman-Coulter). Forward and side scatter measurements were used to differentiate live and dead cells; dead cells were removed from the analyses. Forward scatter height and width measurements were then used to differentiate single cells from doublets and clumps; doublets and clumps were removed from the analyses. Compensation was empirically adjusted by using cells transfected with single-gene constructs (either with the pmCherry-C1 or with the pEGFP-N3 plasmids) and non-transfected cells (See supplementary Fig. 14 for an illustration of the analysis). Data from at least 10,000 cells positive for both mCherry and GFP expression were collected, and further analyses were performed with FlowJo Software (BD). The ratio of green fluorescence relative to red fluorescence were calculated from each individual cell and median, first and third quantiles were derived for 10 000 cells in each experiment. For transfection of the RNA pol II mutants in the stable cell lines, we co-transfected an RFP expressing plasmid in order to select for transfected cells.

Declarations

CONFLICT OF INTEREST

None declared.

AUTHOR CONTRIBUTIONS

Virginie Jouffret: Software and Formal analysis; **Maurane Eisele**: Investigation, Validation, Formal analysis; **Tomas Caetano**: Software and Formal analysis; **Marika Iacovone**: Investigation, Validation, Formal analysis; **Julie Bossuyt**: Investigation and Formal analysis; **Coralie Delpech**: Investigation and Formal analysis; **Marion Aguirrebengoa**: Software and Formal analysis; **Maxime Delmas**: Software and Formal analysis; **Estelle Nicolas**: Supervision, Writing-Review&editing; **Didier Trouche**: Conceptualization, Supervision, Writing-Review&Editing, Funding acquisition; **Lisa Muniz**: Conceptualization, Methodology, Investigation, Validation, Formal analysis, Supervision, Writing-Original draft, Writing-Review&Editing, Project administration, Funding acquisition.

ACKNOWLEDGEMENTS

We thank all members of Didier Trouche's team as well as Sylvain Egloff for helpful discussions; Yvan Canitrot for help with flow cytometry experiments; David Bentley and Nova Fong for RNA pol II speed mutants encoding plasmids; Christophe Lachaud for the pESGRNA plasmid; Kevin Struhl and Catherine Maddox for the MCF10A ER-Src cell line. Thomas Clouaire and Anne-Laure Finoux from the GenSCAN platform (CBI, Toulouse) for advices on libraries preparation for RNAseq experiments. The EMBL Genecore facility for helpful discussions on RNA sequencing.

This work was supported by grants from the Ligue Nationale Contre le Cancer as an "équipe labellisée" and from the ANR PRC H2AZplice to DT, as well as from the Ligue Contre le Cancer 31, the Cancéropole GSO and ANR JCJC 'SPEEDREG' to LM; ME is supported by a studentship from La Fondation pour la Recherche Médicale (ECO202406019147). MI is supported by a studentship from the ANR H2AZplice.

DATA AVAILABILITY

Total RNAseq data from MCF10A-ER-Src cells and TTchem seq data from U2OS cells are deposited in the NCBI Gene Expression Omnibus under accession numbers GSE324304 and GSE342676, respectively.

References

1. Lopez Martinez, D., Svejstrup, J.Q.: Mechanisms of RNA Polymerase II Termination at the 3'-End of Genes. J. Mol. Biol. **437**, 168735 (2025)
2. Eaton, J.D., West, S.: Termination of Transcription by RNA Polymerase II: BOOM! Trends Genet. **36**, 664–675 (2020)
3. Schwalb, B., et al.: TT-seq maps the human transient transcriptome. Science. **352**, 1225–1228 (2016)
4. Shearwin, K.E., Callen, B.P., Egan, J.B.: Transcriptional interference–a crash course. Trends Genet. **21**, 339–345 (2005)
5. Martens, J.A., Laprade, L., Winston, F.: Intergenic transcription is required to repress the Saccharomyces cerevisiae SER3 gene. Nature. **429**, 571–574 (2004)
6. Hainer, S.J., Pruneski, J.A., Mitchell, R.D., Monteverde, R.M., Martens, J.A.: Intergenic transcription causes repression by directing nucleosome assembly. Genes Dev. **25**, 29–40 (2011)
7. Ard, R., Tong, P., Allshire, R.C.: Long non-coding RNA-mediated transcriptional interference of a permease gene confers drug tolerance in fission yeast. Nat. Commun. **5**, 5576 (2014)
8. Ard, R., Allshire, R.C.: Transcription-coupled changes to chromatin underpin gene silencing by transcriptional interference. Nucleic Acids Res. **44**, 10619–10630 (2016)
9. Garg, A., Sanchez, A.M., Shuman, S., Schwer, B.: A long noncoding (lnc)RNA governs expression of the phosphate transporter Pho84 in fission yeast and has cascading effects on the flanking. J. Biol. Chem. **293**, 4456–4467 (2018)
10. Petruk, S., et al.: Transcription of bxd noncoding RNAs promoted by trithorax represses Ubx in cis by transcriptional interference. Cell. **127**, 1209–1221 (2006)
11. Latos, P.A., et al.: Airn transcriptional overlap, but not its lncRNA products, induces imprinted Igf2r silencing. Science. **338**, 1469–1472 (2012)
12. Hollerer, I., et al.: Evidence for an Integrated Gene Repression Mechanism Based on mRNA Isoform Toggling in Human Cells. G3 (Bethesda). **9**, 1045–1053 (2019)
13. Chia, M., et al.: Transcription of a 5' extended mRNA isoform directs dynamic chromatin changes and interference of a downstream promoter. Elife **6**, (2017)
14. Morse, K., Bishop, A.L., Swerdlow, S., Leslie, J.M., Ünal, E.: Swi/Snf chromatin remodeling regulates transcriptional interference and gene repression. Mol. Cell., (2024)
15. Sieriebriennikov, B., et al.: Transcriptional interferences ensure one olfactory receptor per ant neuron. Nature, (2025)
16. Glotzer, G.L., Pastor, P.D.H., Kronauer, D.J.C.: Transcriptional interference gates monogenic odorant receptor expression in ants. Curr. Biol. **35**, 5033–5047e5035 (2025)
17. Boussadia, O., et al.: Transcription of unr (upstream of N-ras) down-modulates N-ras expression in vivo. FEBS Lett. **420**, 20–24 (1997)
18. Muniz, L., et al.: Control of Gene Expression in Senescence through Transcriptional Read-Through of Convergent Protein-Coding Genes. Cell. Rep. **21**, 2433–2446 (2017)

19. Heinz, S., et al.: Transcription Elongation Can Affect Genome 3D Structure. Cell. **174**, 1522–1536e1522 (2018)
20. Rutkowski, A.J., et al.: Widespread disruption of host transcription termination in HSV-1 infection. Nat. Commun. **6**, 7126 (2015)
21. Roth, S.J., Heinz, S., Benner, C.: ARTDeco: automatic readthrough transcription detection. BMC Bioinform. **21**, 214 (2020)
22. Hadar, S., Meller, A., Saida, N., Shalgi, R.: Stress-induced transcriptional readthrough into neighboring genes is linked to intron retention. iScience. **25**, 105543 (2022)
23. Grosso, A.R., et al.: Pervasive transcription read-through promotes aberrant expression of oncogenes and RNA chimeras in renal carcinoma. Elife **4**, (2015)
24. Vilborg, A., et al.: Comparative analysis reveals genomic features of stress-induced transcriptional readthrough. Proc. Natl. Acad. Sci. U S A. **114**, E8362–E8371 (2017)
25. Hennig, T., et al.: HSV-1-induced disruption of transcription termination resembles a cellular stress response but selectively increases chromatin accessibility downstream of genes. PLoS Pathog. **14**, e1006954 (2018)
26. Caldas, P., et al.: Transcription readthrough is prevalent in healthy human tissues and associated with inherent genomic features. Commun. Biol. **7**, 100 (2024)
27. Adhya, S., Gottesman, M.: Promoter occlusion: transcription through a promoter may inhibit its activity. Cell. **29**, 939–944 (1982)
28. Ward, D.F., Murray, N.E.: Convergent transcription in bacteriophage lambda: interference with gene expression. J. Mol. Biol. **133**, 249–266 (1979)
29. Hausler, B., Somerville, R.L.: Interaction in vivo between strong closely spaced constitutive promoters. J. Mol. Biol. **127**, 353–356 (1979)
30. Hirsch, H.A., Iliopoulos, D., Tsichlis, P.N., Struhl, K.: Metformin selectively targets cancer stem cells, and acts together with chemotherapy to block tumor growth and prolong remission. Cancer Res. **69**, 7507–7511 (2009)
31. Lamaa, A., et al.: Integrated analysis of H2A.Z isoforms function reveals a complex interplay in gene regulation. Elife **9**, (2020)
32. Wiesel, Y., Sabath, N., Shalgi, R.: DoGFinder: a software for the discovery and quantification of readthrough transcripts from RNA-seq. BMC Genom. **19**, 597 (2018)
33. Elkon, R., et al.: Myc coordinates transcription and translation to enhance transformation and suppress invasiveness. EMBO Rep. **16**, 1723–1736 (2015)
34. Guo, J.U., Agarwal, V., Guo, H., Bartel, D.P.: Expanded identification and characterization of mammalian circular RNAs. Genome Biol. **15**, 409 (2014)
35. Fong, N., Saldi, T., Sheridan, R.M., Cortazar, M.A., Bentley, D.L.: RNA Pol II Dynamics Modulate Co-transcriptional Chromatin Modification, CTD Phosphorylation, and Transcriptional Direction. Mol. Cell. **66**, 546–557e543 (2017)

36. Wang, Y., Hou, L., Ardehali, M.B., Kingston, R.E., Dynlacht, B.D.: Elongin A regulates transcription in vivo through enhanced RNA polymerase processivity. J. Biol. Chem. **296**, 100170 (2021)
37. Fong, N., et al.: Effects of Transcription Elongation Rate and Xrn2 Exonuclease Activity on RNA Polymerase II Termination Suggest Widespread Kinetic Competition. Mol. Cell. **60**, 256–267 (2015)
38. Fong, N., et al.: Pre-mRNA splicing is facilitated by an optimal RNA polymerase II elongation rate. Genes Dev. **28**, 2663–2676 (2014)
39. Hall, P.M., et al.: Polarity of the CRISPR roadblock to transcription. Nat. Struct. Mol. Biol. **29**, 1217–1227 (2022)
40. Zukher, I., Dujardin, G., Sousa-Luís, R., Proudfoot, N.J.: Elongation roadblocks mediated by dCas9 across human genes modulate transcription and nascent RNA processing. Nat. Struct. Mol. Biol. **30**, 1536–1548 (2023)
41. Dujardin, G., et al.: How slow RNA polymerase II elongation favors alternative exon skipping. Mol. Cell. **54**, 683–690 (2014)
42. Cortazar, M.A., et al.: Control of RNA Pol II Speed by PNUTS-PP1 and Spt5 Dephosphorylation Facilitates Termination by a Sitting Duck Torpedo Mechanism. Mol. Cell. **76**, 896–908e894 (2019)
43. Mei, Y., Cheng, Z., Lu, Y., Wu, S., Chen, X.: Comprehensive resource for transcription readthrough events in healthy human tissues. Sci. Data. **12**, 1176 (2025)
44. Wang, L., Watters, J.W., Ju, X., Lu, G., Liu, S.: Head-on and co-directional RNA polymerase collisions orchestrate bidirectional transcription termination. Mol. Cell. **83**, 1153–1164e1154 (2023)
45. Rodríguez-Martín, B., et al.: ChimPipe: accurate detection of fusion genes and transcription-induced chimeras from RNA-seq data. BMC Genom. **18**, 7 (2017)
46. Varley, K.E., et al.: Recurrent read-through fusion transcripts in breast cancer. Breast Cancer Res. Treat. **146**, 287–297 (2014)
47. Singh, S., et al.: The landscape of chimeric RNAs in non-diseased tissues and cells. Nucleic Acids Res. **48**, 1764–1778 (2020)
48. Vilborg, A., Passarelli, M.C., Yario, T.A., Tycowski, K.T., Steitz, J.A.: Widespread Inducible Transcription Downstream of Human Genes. Mol. Cell. **59**, 449–461 (2015)
49. Nissani, N., Ulitsky, I.: Unique features of transcription termination and initiation at closely spaced tandem human genes. Mol. Syst. Biol. **18**, e10682 (2022)
50. Leng, X., et al.: Organismal benefits of transcription speed control at gene boundaries. EMBO Rep. **21**, e49315 (2020)
51. Yague-Sanz, C., et al.: Nutrient-dependent control of RNA polymerase II elongation rate regulates specific gene expression programs by alternative polyadenylation. Genes Dev., (2020)
52. Shao, R., et al.: Distinct transcription kinetics of pluripotent cell states. Mol. Syst. Biol. **18**, e10407 (2022)
53. Jonkers, I., Kwak, H., Lis, J.T.: Genome-wide dynamics of Pol II elongation and its interplay with promoter proximal pausing, chromatin, and exons. Elife. **3**, e02407 (2014)

54. Veloso, A., et al.: Rate of elongation by RNA polymerase II is associated with specific gene features and epigenetic modifications. Genome Res. **24**, 896–905 (2014)
55. Gregersen, L.H., Mitter, R., Svejstrup, J.Q.: Using TT_{chem}-seq for profiling nascent transcription and measuring transcript elongation. Nat. Protoc. **15**, 604–627 (2020)
56. Danecek, P., et al.: Twelve years of SAMtools and BCFtools. *Gigascience* 10, (2021)
57. Saldi, T., Riemondy, K., Erickson, B., Bentley, D.L.: Alternative RNA structures formed during transcription depend on elongation rate and modify RNA processing. Mol. Cell. **81**, 1789–1801e1785 (2021)
58. Ramírez, F., et al.: deepTools2: a next generation web server for deep-sequencing data analysis. Nucleic Acids Res. **44**, W160–165 (2016)
59. Munoz, I.M., Szyniarowski, P., Toth, R., Rouse, J., Lachaud, C.: Improved genome editing in human cell lines using the CRISPR method. PLoS One. **9**, e109752 (2014)
60. Debnath, J., Muthuswamy, S.K., Brugge, J.S.: Morphogenesis and oncogenesis of MCF-10A mammary epithelial acini grown in three-dimensional basement membrane cultures. Methods. **30**, 256–268 (2003)

Figures

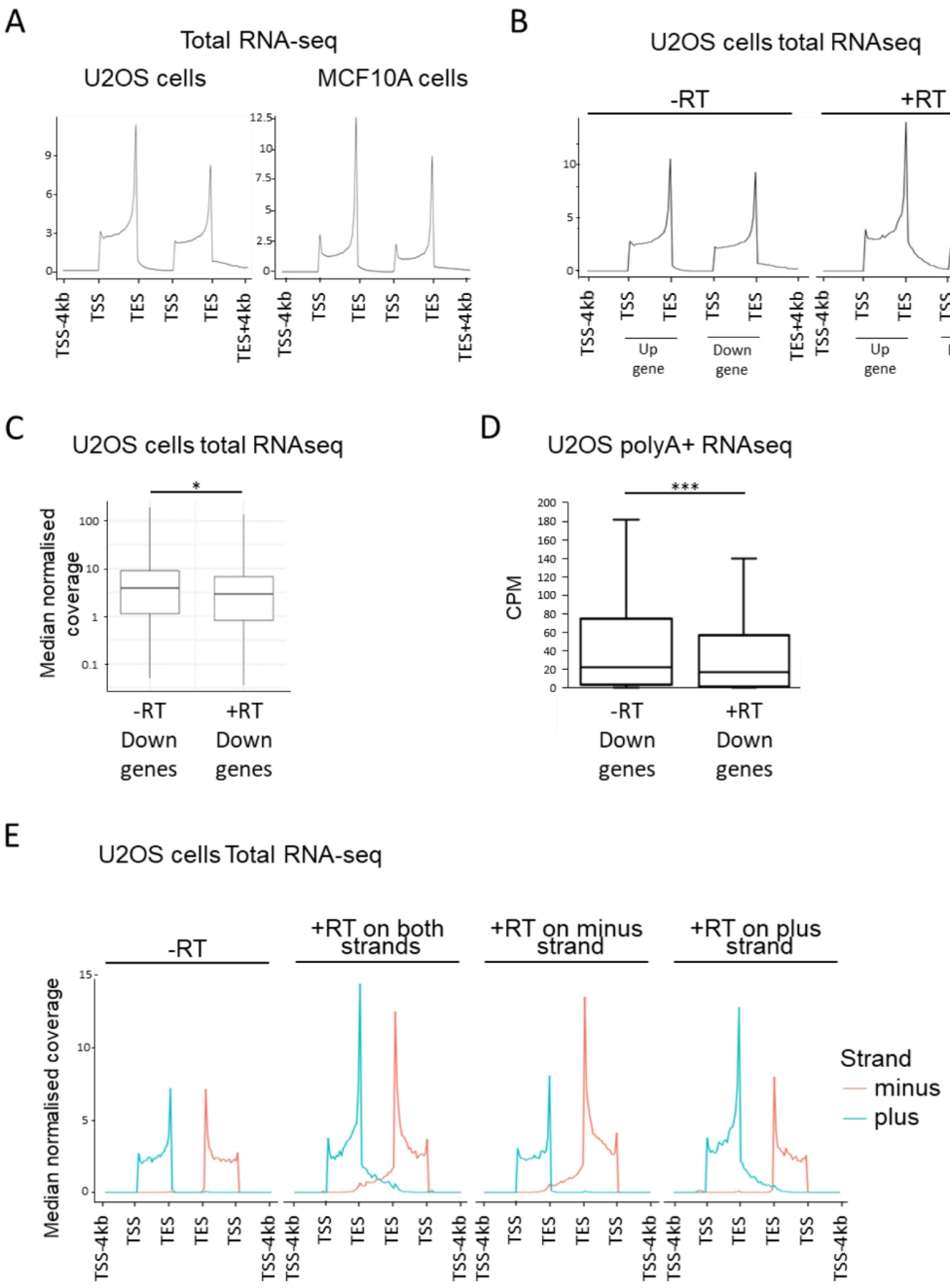


Figure 1

Read-through transcription limits the expression of downstream genes in tandem and convergent gene pairs

(A) Metagene analysis of strand specific total RNA-seq profiles representing the median within the gene population of the normalized coverage in tandem genes and intergenic regions in U2OS and MCF10A

cells. Analyses are shown for one replicate out of two for MCF10A cells and out of four for U2OS cells. Amongst these loci only pairs with a mean of the normalized coverage computed on the gene body >0 and an intergenic region size of more than 100 nt and less than 100kb were considered (4707 loci for U2OS cells and 4071 loci for MCF10A cells). **(B)** Same as in (A) for U2OS cells, except that tandem gene pairs were separated into two classes based on the production or not of read-through transcripts (RT) from the upstream gene detected by the HMM algorithm, as indicated. Analysis of one representative RNA-seq replicate is shown out of four. A minimum threshold on upstream genes expression was applied to -RT gene class in order to avoid biasing our analysis due to a difference in the level of expression of upstream genes in between the -RT and +RT classes (3704 gene pairs in the -RT class; 987 gene pairs in the +RT class). **(C)** Boxplots showing the median, first and third quantiles of the median normalized coverage for downstream genes from the -RT and +RT classes in tandem gene pairs from total RNAseq data from U2OS cells. The -RT and +RT tandem genes classes were selected as in (B). Analyses are shown for one replicate out of four. *p<0.05 by Wilcoxon-Mann-Whitney test and adjusted by the Benjamini-Hochberg procedure. **(D)** Tandem gene pairs were separated into two classes based on the production of read-through transcripts from the upstream gene (-RT and +RT) in published polyA+ RNAseq data from U2OS cells (*33,34*) and the level of expression of downstream genes was determined for each class. The -RT and +RT tandem genes classes were selected as in (B). Boxplots show the median, first and third quantiles of the level of expression of these genes expressed in CPM. ***p<0.01 by Wilcoxon-Mann-Whitney test and adjusted by the Benjamini-Hochberg procedure. **(E)** Same as in (B) except that convergent genes' loci were analysed and separated based on the strand on which the gene producing read-through transcripts is located (1044, 356, 575, and 555 loci in the -RT, +RT on both strands, +RT on the minus strand, and +RT on the plus strand classes, respectively).

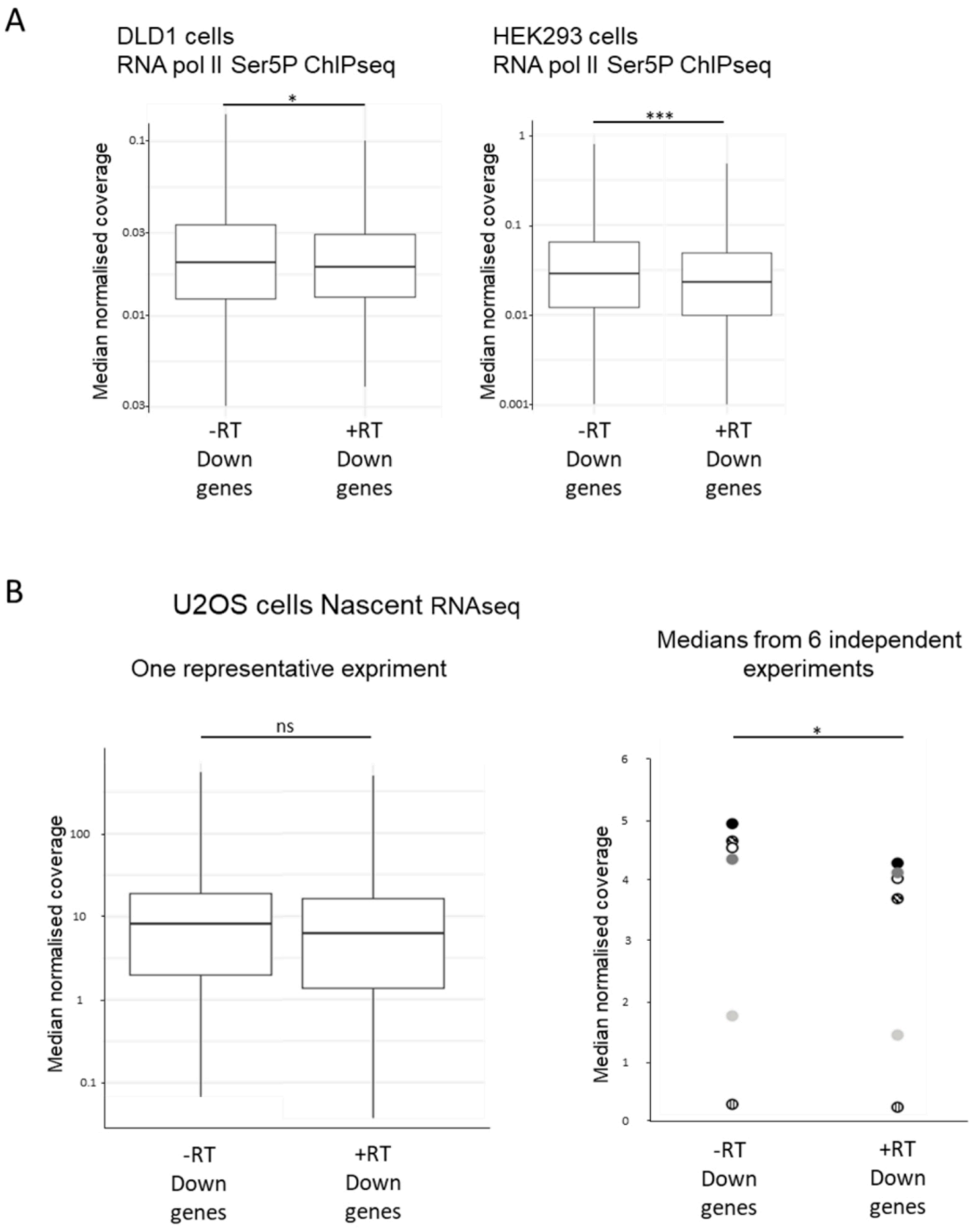


Figure 2

Read-through expression limits transcription of downstream genes in tandem gene pairs

(A) Analysis of the median normalized coverage of S5P RNA pol II ChIPseq data from DLD1 (*36*) and HEK293 cells (*35*). Tandem gene pairs were separated into two classes based on the production of read-through transcripts (RT) from the upstream gene or not as indicated. Amongst these loci only pairs with

a mean of the normalized coverage computed on the gene body >0 and an intergenic region size of more than 100nt and less than 100kb were considered. The level of RNA pol II Ser5P detected on downstream genes was determined for each class. Boxplots show the median, first and third quantiles of the median normalized coverage of the ChIPseq on these genes. **(B)** Tandem gene pairs were separated into two classes based on the production of read-through transcripts from the upstream gene or not (-RT and +RT) in TTchemseq data from U2OS cells and the level of expression of downstream genes was determined for each class. Left panel: Analyses are shown for one replicate out of six. Boxplots show the median, first and third quantiles of the median normalized coverage for these genes. ns: non-significant. Right panel: The dot plots show the medians of the median normalized coverage for downstream genes in the -RT and +RT classes from six independent experiments, with individual experiments identified by circles of different colors. *p<0.05 by paired Welch test.

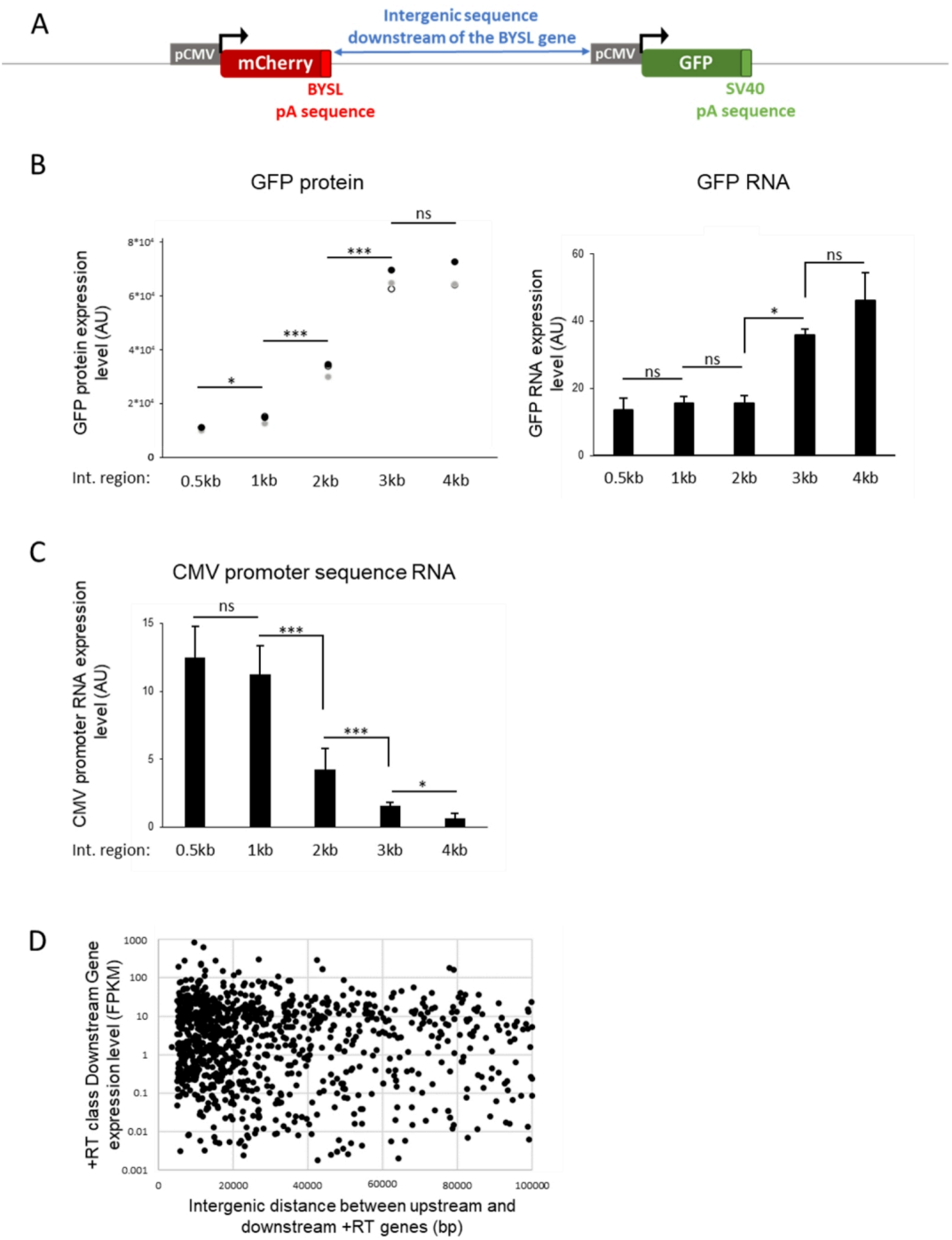


Figure 3

The distance travelled by RNA pol II influences transcriptional interference on the downstream gene

(A) Schematic representation of the reporter system used to analyse transcriptional interference. The mCherry and GFP coding sequences were cloned downstream of two identical CMV promoters and separated by genomic sequences of different lengths beginning at a position located 109bp upstream of

the polyA site of the human BYSL gene. **(B)** Left panel: GFP protein expression level measured by flow cytometry in U2OS cells transfected with the different reporter systems. The dot plots show the medians for 10 000 cells from three independent experiments, with individual experiments identified by circles of different colors. ns: non-significant, *p<0.05, ***p<0.01 by paired Welch test. Right panel: GFP RNA expression level measured by RT-qPCR in U2OS cells transfected with the different reporter systems, normalized by the level of expression of the puromycine resistance gene which is expressed from the reporter plasmid in order to account for transfection efficiency differences in between samples. The histogram shows the means and standard deviation from three independent experiments. *p<0.05, ns: non-significant by Student's T-test. **(C)** Expression level of CMV promoter RNAs measured by RT-qPCR in U2OS cells transfected with the different reporter systems normalized by the level of expression of puromycin resistance gene expressed from the reporter plasmids. Means and standard deviation from three independent experiments are shown. *p<0.05, ***p<0.01, ns: non-significant by Student's T-test. **(D)** Dot plot representing the level of expression (FPKM) of downstream genes from the +RT tandem genes' class according to the intergenic distance separating them from the upstream gene expressed in bp.

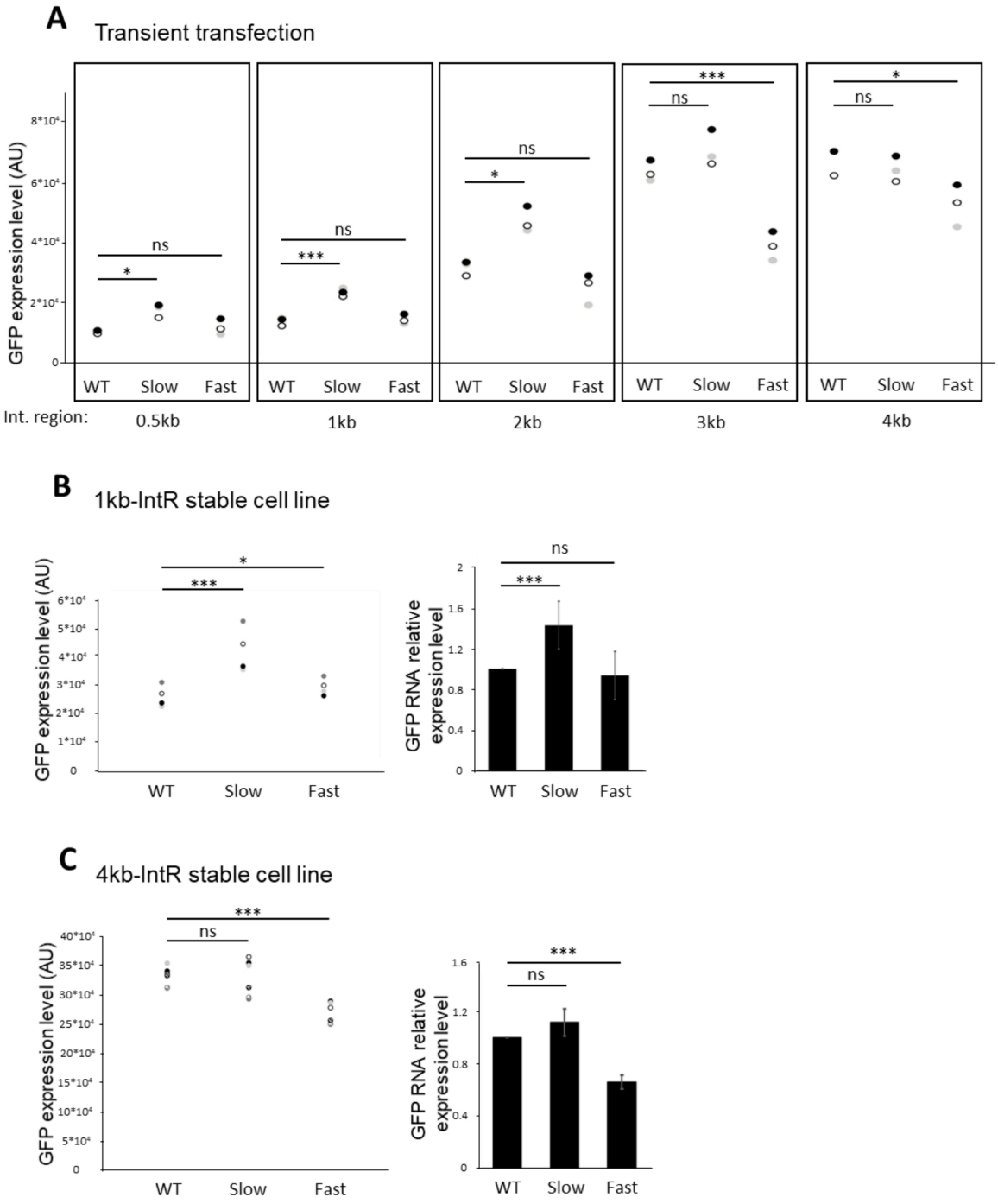


Figure 4

RNA pol II speed controls the extent of transcriptional interference

(A) GFP protein expression levels were measured by flow cytometry in U2OS cells co-transfected with plasmids encoding either the 0.5kb-IntR, 1kb-IntR, 2kb-IntR, 3kb-IntR or 4kb-IntR reporter systems together with the WT, Slow or Fast RNA pol II mutants expression vectors and treated with alpha-

amanitin for 48 hours. Dot plots show the median for 10 000 cells in each condition, from three independent experiments identified by circles of different colors. *p<0.05, ***p<0.01, ns: non-significant by Welch t-test and adjusted by the Benjamini-Hochberg procedure. **(B)** Left panel: The 1kb-IntR reporter cell line was transfected with plasmids encoding the WT, Slow or Fast RNA pol II mutants and treated with alpha-amanitin for 48 hours. The GFP protein expression levels were measured by flow cytometry. Dot plots show the median for 10 000 cells in each condition, from four independent experiments identified by circles of different colors. ***p<0.01, *p<0.05 by Welch t-test. Right panel: GFP RNAs expression levels were measured by RT-qPCR and normalized relative to GAPDH mRNA expression level. The means and standard deviations from four independent experiments are shown. ***p<0.01, ns: non-significant by Student's T-test. **(C)** Same as in (B) except that the 4kb-IntR reporter cell line was used. Left panel: Dot plots show the median for 10 000 cells in each condition, from six independent experiments identified by circles of different colors.

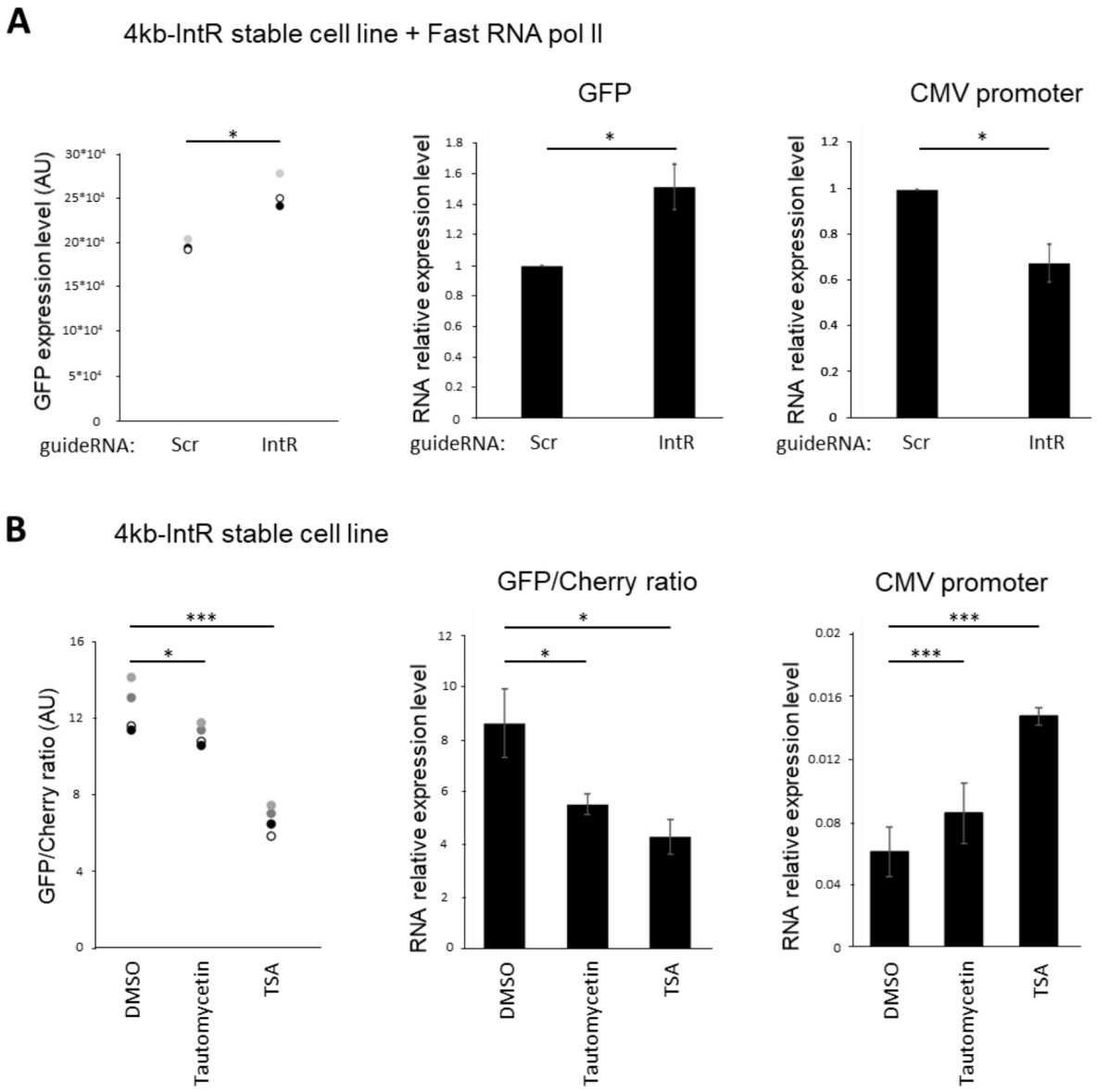


Figure 5

Perturbations of the endogenous RNA pol II speed influence transcriptional interference

(A) The 4kb-IntR reporter cell line was co-transfected with plasmids encoding the Fast RNA pol II mutant, a dCas9 protein, and either a control guide RNA (Scr) or a guide RNA targeting the non-transcribed strand in the intergenic region between mCherry and GFP (IntR). Left panel: GFP protein expression level was measured by flow cytometry for each individual cell in the population for 10 000 cells in each condition in each experiment. Dot plots show the medians from 3 independent experiments identified by circles of different colors. *p<0.05 by Student's t-test. Middle and Right panels: GFP and CMV promoter RNAs expression levels were measured by RT-qPCR and normalized relative to GAPDH mRNA expression

level. The means and standard deviations from three independent experiments are shown. *p<0.05, by Student's t-test. **(B)** Left panel: GFP and mCherry proteins expression levels were measured by flow cytometry in the 4kb-IntR cell line treated with either the histone deacetylase inhibitor trichostatin A (TSA) at 3.3 µM, an inhibitor of the PP1 phosphatase Tautomycetin at 500nM or DMSO as a control. The ratio of GFP over mCherry expression was calculated for each individual cell in the population. Dot plots show the medians for 10 000 cells from four independent experiments identified by circles of different colors. *p<0.05, ***p<0.01 by Welch t-test. Middle and Right panels: GFP/mcherry ratio and CMV promoter RNAs expression levels were measured by RT-qPCR and normalized relative to GAPDH mRNA expression level. The means and standard deviations from five independent experiments are shown. *p<0.05, ***p<0.01 by Student's t-test.

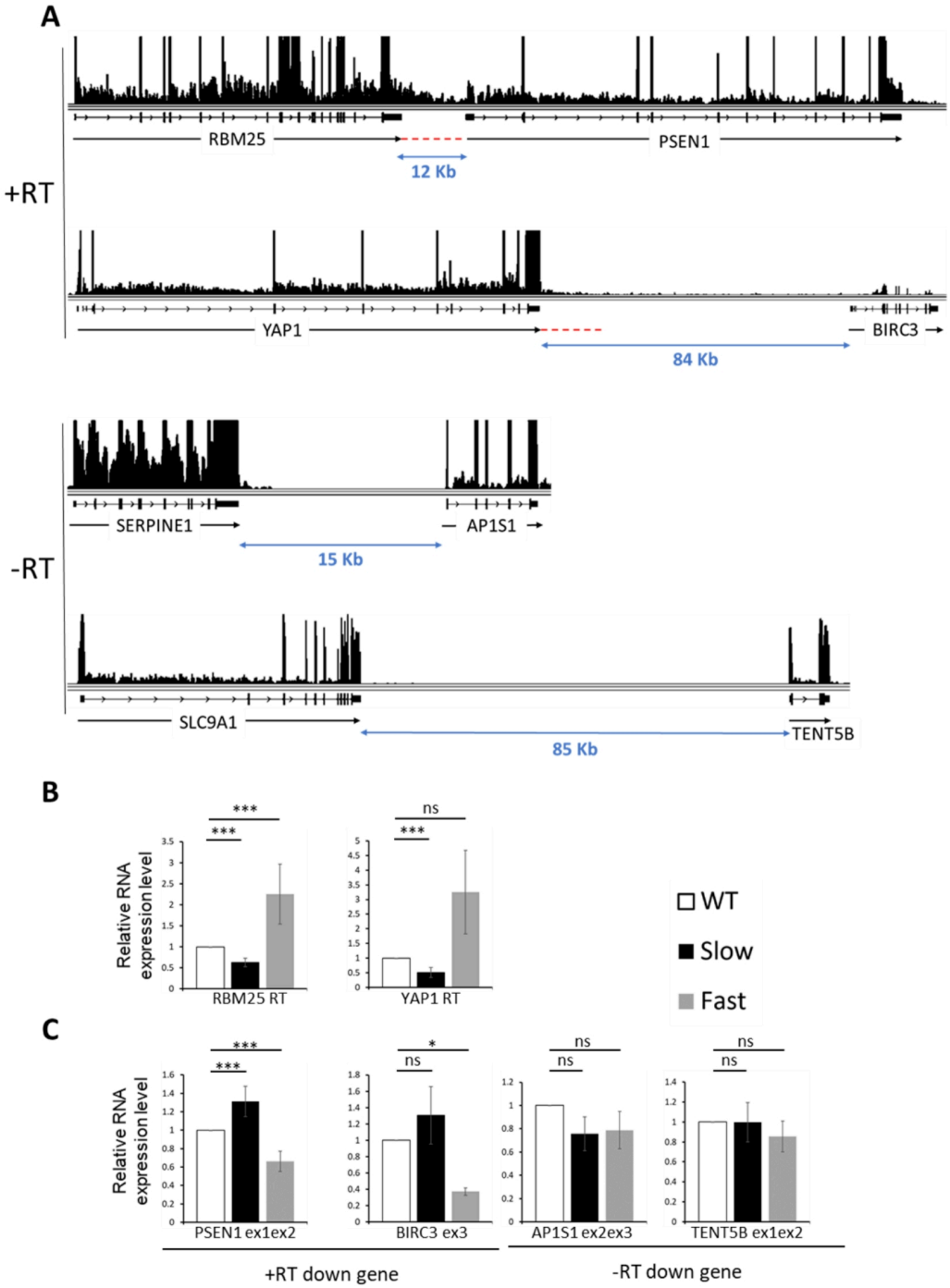


Figure 6

RNA pol II speed influences the level of transcriptional interference on endogenous genes

(A) IGV genome browser views of total RNAseq data from U2OS cells for the RBM25-PSEN1 and YAP1-BIRC3 tandem gene pairs which are part of the +RT class and SERPINE1-AP1S1 and SLC9A1-TENT5B tandem gene pairs which are part of the -RT class. The black arrows indicate the orientation of the

genes. The red dotted line indicates the read-through transcript detected by the HMM algorithm for the RBM25 and YAP1 genes. The blue double arrows indicate the position and length of the intergenic region. **(B)** U2OS cells were transfected with plasmids encoding the WT, Slow or Fast RNA pol II mutants and treated with alpha-amanitin for 48 hours. The expression level of read-through transcripts was analysed by RT-qPCR for the +RT genes; the amplicon is located 11.8kb downstream of RBM25 polyA site and 0.7kb upstream of PSEN1 TSS for the RBM25-PSEN1 pair and 9.5kb downstream of YAP1 polyA site for the YAP1-BIRC3 pair. The means and standard deviations from eight independent experiments are shown. ***p<0.01, ns: non-significant by Student's t-test. **(C)** Same as in (B) except that the level of expression of downstream genes for each tandem gene pair was measured by RT-qPCR and normalized to GAPDH mRNA expression level. The means and standard deviations from eight independent experiments are shown. *p<0.05, ***p<0.01, ns: non-significant by Student's t-test.

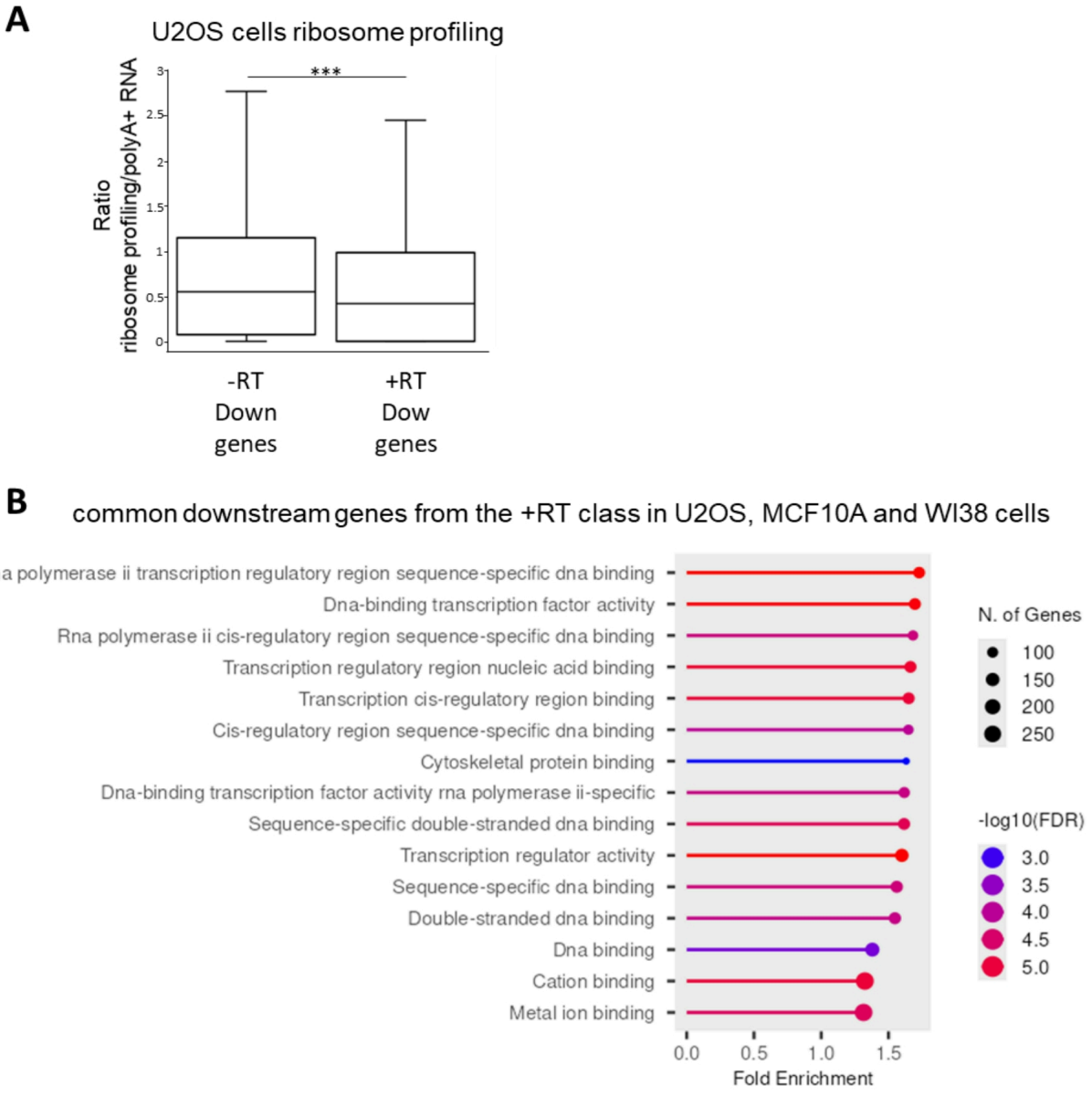


Figure 7

RNA pol II speed influences the level of transcriptional interference on endogenous genes

(A) IGV genome browser views of total RNAseq data from U2OS cells for the RBM25-PSEN1 and YAP1-BIRC3 tandem gene pairs which are part of the +RT class and SERPINE1-AP1S1 and SLC9A1-TENT5B tandem gene pairs which are part of the -RT class. The black arrows indicate the orientation of the genes. The red dotted line indicates the read-through transcript detected by the HMM algorithm for the RBM25 and YAP1 genes. The blue double arrows indicate the position and length of the intergenic region. **(B)** U2OS cells were transfected with plasmids encoding the WT, Slow or Fast RNA pol II mutants and treated with alpha-amanitin for 48 hours. The expression level of read-through transcripts was analysed by RT-qPCR for the +RT genes; the amplicon is located 11.8kb downstream of RBM25 polyA site and 0.7kb upstream of PSEN1 TSS for the RBM25-PSEN1 pair and 9.5kb downstream of YAP1 polyA site for the YAP1-BIRC3 pair. The means and standard deviations from eight independent experiments are shown. ***p<0.01, ns: non-significant by Student's t-test. **(C)** Same as in (B) except that the level of expression of downstream genes for each tandem gene pair was measured by RT-qPCR and normalized to GAPDH mRNA expression level. The means and standard deviations from eight independent experiments are shown. *p<0.05, ***p<0.01, ns: non-significant by Student's t-test.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplfiguresJouffretEiseleetal.docx