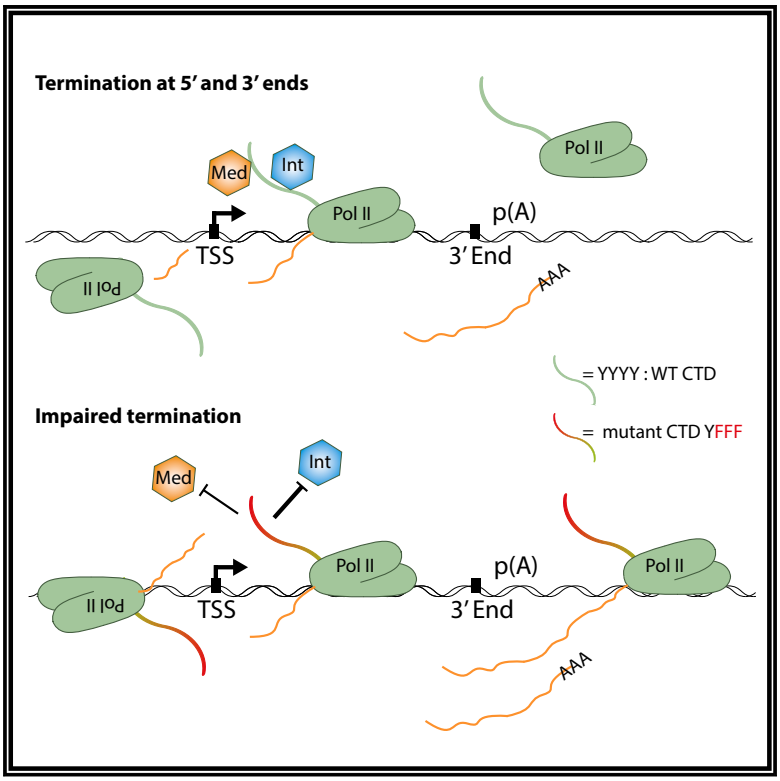
Article

Molecular Cell

Tyrosine-1 of RNA Polymerase II CTD Controls Global Termination of Gene Transcription in Mammals

Graphical Abstract



Highlights

- Mammalian Pol II CTD Tyrosine1 is required for transcription termination control
- Tyrosine1 mutations result in massive read-through transcription at gene ends
- Tyrosine1 mutations impair Pol II association with Mediator and Integrator
- Tyrosine1 are involved in the maturation and/or stability of non-polyadenylated RNAs

Authors

Nilay Shah, Muhammad Ahmad Maqbool, Yousra Yahia, ..., Axel Imhof, Dirk Eick, Jean-Christophe Andrau

Correspondence

eick@helmholtz-muenchen.de (D.E.), jean-christophe.andrau@igmm.cnrs.fr (J.-C.A.)

In Brief

Transcription of eukaryotic genes requires an efficient termination to avoid pervasive transcript synthesis. Here, Shah and Maqbool et al. show that tyrosine residues of RNA polymerase II CTD are essential for termination and recruitment of the Mediator and Integrator complexes. A massive read-through phenotype is observed when these residues are mutated.

Shah et al., 2018, Molecular Cell 69, 48–61
January 4, 2018
https://doi.org/10.1016/j.molcel.2017.12.009

CellPress

CellPress

Molecular Cell
Article

Tyrosine-1 of RNA Polymerase II CTD Controls Global Termination of Gene Transcription in Mammals

Nilay Shah,[1,5] Muhammad Ahmad Maqbool,[2,5] Yousra Yahia,[2] Amal Zine El Aabidine,[2] Cyril Esnault,[2] Ignasi Forné,[3] Tim-Michael Decker,[1] David Martin,[2] Roland Schüller,[1] Stefan Krebs,[4] Helmut Blum,[4] Axel Imhof,[3] Dirk Eick,[1,*] and Jean-Christophe Andrau[2,6,*]

[1]Department of Molecular Epigenetics, Helmholtz Center Munich and Center for Integrated Protein Science Munich (CIPSM), Marchioninistrasse 25, 81377 Munich, Germany
[2]Institut de Génétique Moléculaire de Montpellier (IGMM), Univ Montpellier, CNRS, Montpellier, France
[3]Biomedical Center Munich, ZFP, Großhaderner Strasse 9, 82152 Planegg-Martinsried, Germany
[4]Laboratory for Functional Genome Analysis, Gene Center, Ludwig-Maximilians-Universität, Munich, Germany
[5]These authors contributed equally
[6]Lead Contact
*Correspondence: eick@helmholtz-muenchen.de (D.E.), jean-christophe.andrau@igmm.cnrs.fr (J.-C.A.)
https://doi.org/10.1016/j.molcel.2017.12.009

SUMMARY

The carboxy-terminal domain (CTD) of RNA polymerase (Pol) II is composed of a repetition of YSPTSPS heptads and functions as a loading platform for protein complexes that regulate transcription, splicing, and maturation of RNAs. Here, we studied mammalian CTD mutants to analyze the function of tyrosine1 residues in the transcription cycle. Mutation of 3/4 of the tyrosine residues (YFFF mutant) resulted in a massive read-through transcription phenotype in the antisense direction of promoters as well as in the 3′ direction several hundred kilobases downstream of genes. The YFFF mutant shows reduced Pol II at promoter-proximal pause sites, a loss of interaction with the Mediator and Integrator complexes, and impaired recruitment of these complexes to chromatin. Consistent with these observations, Pol II loading at enhancers and maturation of snRNAs are altered in the YFFF context genome-wide. We conclude that tyrosine1 residues of the CTD control termination of transcription by Pol II.

INTRODUCTION

The control of transcription requires RNA polymerase (Pol) II recruitment at promoter, transcription initiation, and transition to processive elongation. It also requires a proper control of transcription termination (Proudfoot, 2016). Despite many efforts during the past years to understand this process *in vivo*, it remains poorly understood. Transcription termination by Pol II generally succeeds polyadenylation at 3′ ends of genes and can occur up to several kilobases after the annotated 3′ ends. Recent works have involved proteins or protein complexes in this process, such as the cleavage and polyadenylation complex and the histone methyl-transferase SetD2 (Grosso et al., 2015; Nojima et al., 2015). Termination also occurs at 5′ ends of genes. This process concerns a large fraction of mammalian promoters in which the pausing of Pol II and divergent transcription are observed (Core et al., 2008; Seila et al., 2008). Longer upstream antisense (AS) non-coding transcripts can also be observed at many promoters in normal cells and accumulate to high levels after exosome inhibition (Lepoivre et al., 2013; Preker et al., 2008; Schlackow et al., 2017). Current models propose that termination around promoters also requires polyadenylation, a process that would be partially repressed in the sense, but not in the antisense, orientation of the genes by the presence of U1 snRNP recognition sites (Almada et al., 2013; Ntini et al., 2013).

Among the questions left open is that of the determinants targeted within the Pol II enzyme allowing termination in living cells. Although mutations in catalytic subunits were characterized in Pol III active sites that impair transcription termination (Iben et al., 2011; Shaaban et al., 1995), little is known about Pol II, despite reports of slow/fast Pol II mutants displaying impaired termination at a subset of genes (Fong et al., 2015; Hazelbaker et al., 2013). In contrast to Pol I and Pol III, Pol II produces transcripts of widely varying sizes and types, including polyadenylated, non-polyadenylated, coding, and non-coding transcripts with various functions and, thus, different modes of regulation. As a consequence, Pol II activity is tightly controlled through the action of many proteins or protein complexes that can act at all steps of transcription, including recruitment, initiation, pausing, pause release, processive elongation, and termination.

The carboxy-terminal domain (CTD) of Pol II's largest subunit, Rpb1, is an essential platform for recruitment of factors controlling transcriptional and post-transcriptional events (Eick and Geyer, 2013; McCracken et al., 1997). The CTD is evolutionarily conserved and consists of repetitions of

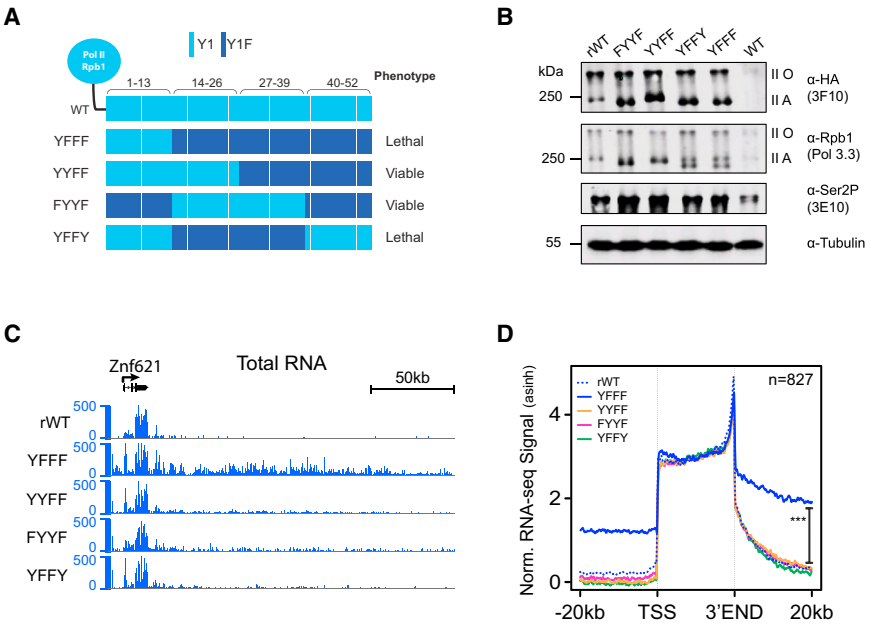


Figure 1. Screening of Human CTD Tyrosine Mutants

(A) Schematic representation of Pol II CTD tyrosine mutants. WT and mutant heptads are represented in light and dark blue, respectively.

(B) Western blot of rWT and CTD mutants following 24-hr induction and 48-hr α-amanitin treatment of the cells (72-hr induction).

(C) Example of RT phenotype at the 3′ end of the *Znf621* gene.

(D) Average metagene profile of total sense RNA-seq signal (inverse hyperbolic sine [asinh] transformed) over the gene bodies and 20-kb upstream and downstream regions. All profiles were normalized so that signals are equivalent on gene bodies (see STAR Methods). $***p < 2 \times 10^{-16}$ (2-sided Wilcoxon test) between rWT and YFFF. See also Figure S1.

heptads ($Y_1S_2P_3T_4S_5P_6S_7$) that are phosphorylated in the transcription cycle. Phosphorylations of serine 2 and serine 5 (Ser2P and Ser5P) residues are the most studied and represent strong hallmarks of early transcription and processive elongation, respectively. The more recently characterized Ser7P and Thr4P were proposed to be associated with small nuclear RNA (snRNA) or histone gene transcription and transcription termination (Chapman et al., 2007; Egloff et al., 2007; Harlen et al., 2016; Hintermair et al., 2012; Hsin et al., 2011).

We and others have recently described that phosphorylation of Tyrosine1 (Tyr1) in metazoans occurs at promoters (Descostes et al., 2014; Hsin et al., 2014) and that, in mammals, Tyr1P is also found at enhancer locations. Chromatin immunoprecipitation sequencing (ChIP-seq) signals for Tyr1P were also observed to a lesser extent at 3′ ends. Overall, the mammalian Tyr1P genomic locations were quite distinct from the ones described in yeast, where enrichments were essentially found over gene bodies and proposed to prevent early termination (Mayer et al., 2012). However, we were previously unable to describe the functional significance of Tyr1 residues due to the lack of stable mutants, as mutations of all Tyr1 residues of the CTD resulted in the degradation of Rpb1 (Descostes et al., 2014). To circumvent this problem, we have generated novel mutations in the CTD and focused our analyses on a mutant, YFFF, in which Tyr1 residues are replaced by Phe in the last 3/4 of the CTD repeats. This mutant reveals a role of Tyr1 residues in the control of the termination of 5′ antisense and 3′ sense transcripts. In the YFFF mutant, a massive transcription read-through (RT) is observed, accompanied by reduced Pol II at the promoter-proximal pause, apparent transcriptional interference, snRNA maturation defect, and decrease of Pol II accumulation at active enhancers. Further proteomic characterization of the YFFF mutant showed that tyrosine mutations resulted in the loss of Pol II interaction with Mediator (Med) and Integrator (Int) complexes, suggesting that they might be involved in the pause/termination processes. Finally, both Med and Int also show impaired DNA recruitment as revealed by ChIP experiments.

RESULTS

Phenotypes of CTD Tyrosine Mutants

We previously investigated the function of Tyr1P residues in the mammalian CTD by genome-wide location analysis (ChIP-seq) and by generating mutations in the CTD, replacing all Tyr1 residues of CTD heptads into phenylalanine (Descostes et al., 2014). These mutations resulted in a lethal phenotype and CTD degradation in Rpb1, restricting further functional investigation. To circumvent this problem, we designed four new CTD mutants (Figure 1A) in which only Tyr1 residues of 2 or 3 quarters of the heptads were mutated to Phe residues. The control used in our experiments contains the wild-type (WT) CTD sequence, including the non-canonical repeats, and is designated as rWT. All mutants, as well as rWT, contain an α-amanitin resistance mutation that allows to express a recombinant Rpb1, while the endogenous Rpb1 is suppressed as described previously (Bartolomei et al., 1988; Meininghaus et al., 2000).

After induction of the mutants and rWT control cells, endogenous Rpb1 was shut down by α-amanitin treatment. We then analyzed the growth phenotype and the stability of the mutants. Mutants with half of the repeats mutated were found either lethal (YFFY) or viable and proliferated (YYFF and FYYF) for 5 to 10 days after the addition of α-amanitin to the medium, suggesting that the position of the heptads within the CTD is important for tyrosine function (Figures S1A and S1B). Mutation of last three quarters of the repeats (YFFF) also resulted in a lethal phenotype. Despite their variable phenotypes, all Rpb1 mutants were stable at the protein level with comparable amounts of hyper- (IIO) versus hypo-phosphorylated (IIA) forms of Rpb1 (Figure 1B), as well as a comparable level of Ser2P, suggesting that they are competent for elongation and allowing us to pursue

functional study on the mutants. We also monitored the expression of the various phospho-isoforms of the CTD and found comparable levels of Ser2P, Ser5P, Thr4P, and Ser7P (Figure S1C). At the time of sample collection for further experiments, all mutant cells displayed around 80% viability.

We next assessed how transcriptomes of mutants were affected by performing RNA sequencing (RNA-seq) experiments after induction of recombinant Pol II and inhibition of endogenous Pol II with α-amanitin treatment (Figure S1A). In global differential expression (DEseq) analysis, we found a large dysregulation essentially in the YFFF mutant, with many genes downregulated (48) and upregulated (810) (Figure S1D). However, Gene Ontology analyses did not reveal specific functional categories that were lost or enriched in the mutant (data not shown). Rather than an effect at specific categories of genes, our observations pointed to a global effect characterized by a 3′ RT phenotype visible weakly in mutants YYFF, FYYF, and YFFY but strongly pronounced in the YFFF mutant (Figures 1C, 1D, S1E, and S1F). The extent of the observed RT in the YFFF mutant appeared extreme, spanning from several kilobases up to hundreds of kilobases from the annotated 3′ ends, suggesting a global pervasive phenotype. The phenomenon of 3′ RT has been reported for WT Pol II before (Proudfoot, 2016) and has been described with a more amplified phenotype after knockdown of Setd2, Xrn2, CPSF, or WDR82 proteins (Austenaa et al., 2015; Fong et al., 2015; Grosso et al., 2015; Nojima et al., 2015). All these factors are known to interact with CTD and function in the control of RNA elongation/termination. Interestingly, YFFF mutation has little effect on the binding of these factors to CTD, while the interaction with other factors and cellular complexes is fully abolished (mass spectrometry data are discussed later). In sum, the phenotype of the YFFF mutant suggests a strong functional link between Tyr1 in the CTD and the control of termination.

Tyrosine Mutations Cause a Massive RT at 5′ and 3′ Ends of Genes

To examine the consequences of the YFFF mutations in more detail and to strengthen our initial observations, we undertook further total RNA-seq experiments in which we improved the signal-to-noise ratio in intergenic regions (see STAR Methods) as exemplified in Figure S2A and quantified genome-wide in Figure S2B. Using this procedure, we confirmed a massive 3′ RT phenotype in the YFFF mutant and also observed an RT for 5′ antisense transcription. An example for both phenotypes is shown for the *PDCD6IP* gene in Figure 2A. 5′ antisense transcription is a hallmark of mammalian genes (Core et al., 2008; Seila et al., 2008) that occurs roughly at half of the promoters (Fenouil et al., 2012). To consolidate this observation at the genome-wide scale, we performed RNA-seq composite average metagene profiles for protein-coding genes by rescaling rWT and YFFF RNA signals at the same levels over the gene bodies (Figures 2B and S2C) to better visualize the RT phenotype. This demonstrated a clear RT effect at the 3′ ends of genes in the sense direction and the 5′ ends of genes in the antisense direction. This effect was also clearly visible and significant without normalization of signals at gene body (Figure S2D). We confirmed this independently by plotting the transcript densities over the gene bodies and 20 kb upstream of 5′ ends and downstream of 3′ ends (Figure 2C). The transcriptome (gene bodies; Figure 2C, middle panel) shows a typical bimodal distribution representing lowly and moderately/highly expressed genes. The YFFF mutant displays more low expression values (first Gaussian) and less moderate/high values (second Gaussian) as compared to the rWT. The distributions of the 5′ antisense and 3′ sense signals of the 20-kb regions surrounding the gene bodies indicate an inverse trend with more signal for the YFFF mutant. The quantification of upstream antisense and downstream sense RT indices in the rWT and YFFF mutant is shown in Figure 2D. Finally, a larger chromosomal view (Figure 2E) further supports the genome-wide effect of the RT phenotype. Altogether, our data suggest a strong termination defect in mutant YFFF occurring at both ends of genes in sense (downstream) and antisense (upstream) orientations.

We next asked whether the termination defect was specific to the tyrosine mutations of the CTD. To this end, we built a serine2-to-alanine mutant in which 3/4 of the distal CTD repeats are mutated (S2AAA) in a manner similar to that of the YFFF mutant. After shutdown of the endogenous Pol II, the S2AAA mutant also showed a lethal phenotype but no significant 3′ RT and only a slight increase in 5′ antisense transcription (Figure S2E; data not shown) at a few genomic locations. Thus, the observed RT phenotype is specific to the mutation of tyrosine residues in the CTD.

Previous works proposed that Pol II loading at the 5′ end of genes could influence termination at 3′ ends (Nagaike et al., 2011; Pinto et al., 2011). We sought to address this question in the context of the YFFF mutations and, more specifically, whether a marked 3′ RT is linked to increase in 5′ antisense transcription and vice versa. We ranked the genes for decreasing ratio of RNA-seq signal downstream of 3′ ends in YFFF mutant versus rWT, split them into 4 groups (A to D) from the highest to the lowest RT effect at the 3′ end, and plotted the 5′ antisense RNA signal correspondingly (Figure S2F; Table S8). Our analysis reveals that a high RT transcription index at the 3′ end of genes in groups A and B correlated with a high RT transcription index for divergent transcription at the 5′ ends of genes. Lower levels of 3′ end RT transcription in groups C and D were paralleled by lower levels of RT antisense transcription at 5′ ends of genes. Similar analyses were performed for larger intervals (20 and 50 kb downstream of 3′ ends) and also by ranking the genes for decreasing ratio of 5′ antisense RT transcription (data not shown) and indicated a link between 5′ antisense transcription and 3′ RT phenotype in the YFFF mutant. Further investigations however, did not allow us to directly correlate the 5′ to 3′ RT levels in rWT or YFFF cells.

Altogether, our investigations support a model in which the antisense RT transcription at 5′ ends of genes and the 3′-end RT are linked in the context of the YFFF mutant. They also indicate that at least half of the genes do present a significant RT phenotype. In summary, our RNA-seq experiments strongly suggest that Tyr1 mutations of the CTD result in a massive and specific termination defect that occurs both for 5′ antisense transcripts and sense transcripts at 3′ ends of the genes.

The YFFF Mutations Result in Transcriptional Activation of Downstream Genes and Transcriptional Interference

Because the YFFF mutant displayed an apparently pervasive transcription phenotype, we asked whether this could result in

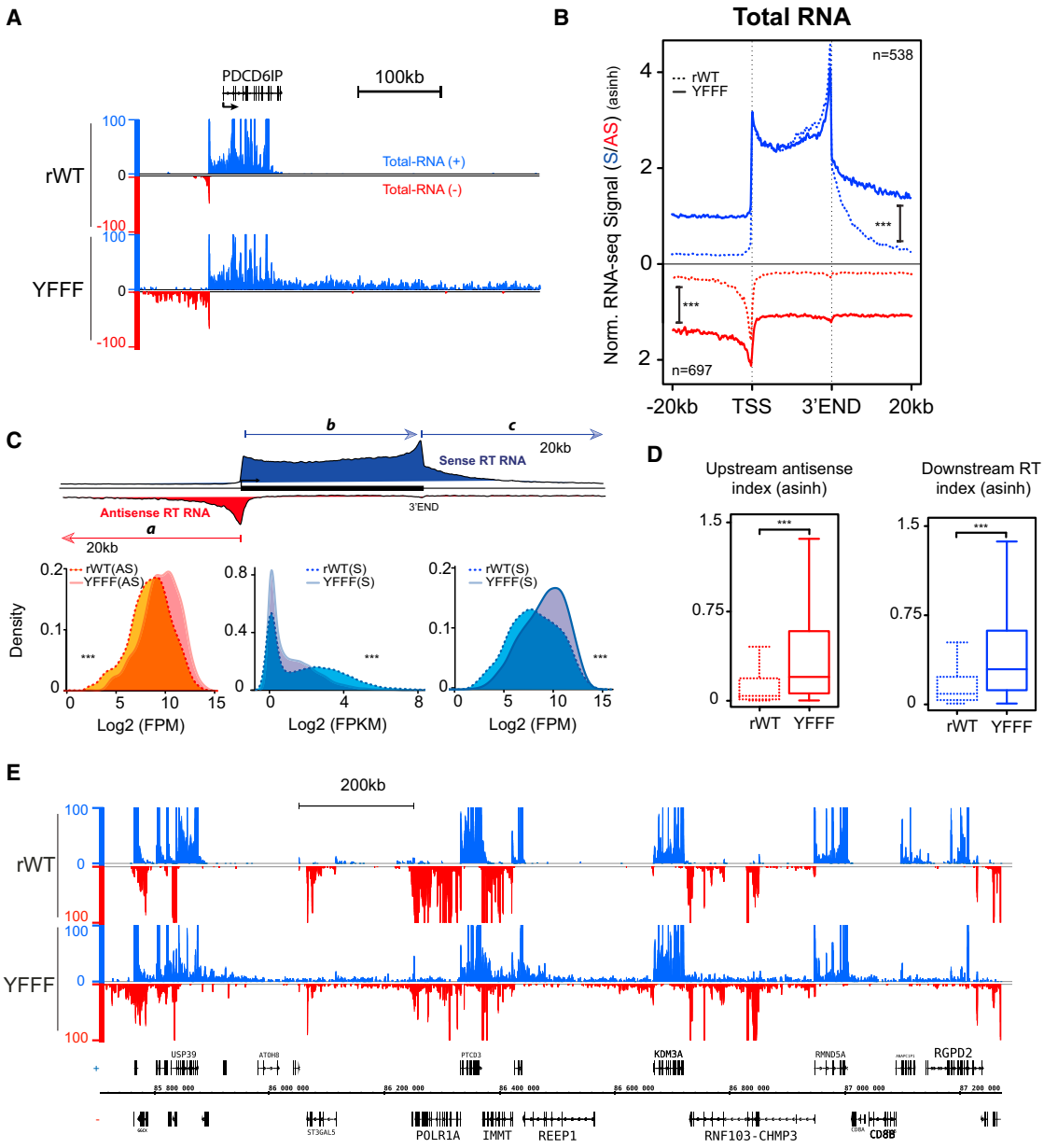


Figure 2. YFFF Mutations Cause a Massive RT Both at the 3′ and 5′ (Antisense) Ends of Genes

(A) Example of RNA-seq signal (y axis) for a coding gene showing the 5′ (antisense) and 3′ (S) RT that extends to at least 100 kb upstream of the 5′ end and 300 kb downstream of the 3′ end.

(B) Average metagene profile of total RNA-seq signal (asinh) in sense (blue) and antisense (red) orientation of the gene bodies and 20-kb upstream and downstream regions.

(C) Density plots of antisense RNA-seq signal in the 20-kb region upstream (a) or downstream (c) of the genes (fragments per million nucleotides; FPM) or sense signal on gene body (fragments per kilobase of transcript per million mapped reads; FPKM in b) in rWT and YFFF cells. Selected regions were excluding genes <2 kb and/or having other genes within 20 kb. Regions concern 1,160 upstream antisense areas, 3,999 gene bodies, and 1,263 downstream areas of the genome. All pairs of distribution are significantly different with a p value $< 2 \times 10^{-16}$ (2 sided Wilcoxon test).

(D) Boxplot of upstream antisense indices (left) and downstream RT indices (right). Units are asinh transformed.

(E) Chromosome 2 snapshot of total RNA-seq data illustrating the generality of the YFFF RT phenotype. See also Figure S2.

transcriptional interference or transcription of previously silent genes due to RT. Visual inspection of our data reveals many examples in which RT transcription of one gene resulted in increased transcription of the downstream gene (see *PPFIA4* gene in Figure S2G). In this case, we could exclude that signals originated from new initiation, as no H3K4me3 (or H3K27ac) signal was observed in the intergenic regions or at the promoter of the downstream gene. Conversely, when two genes were

oriented head to head, we found many examples of apparent interference of the RT with transcription of adjacent genes (see *ST14* gene in Figure S2H). However, we did not observe the loss of H3K4me3 marks at adjacent promoters, possibly because erasing of this histone modification is not very dynamic. In an attempt to quantify global interference, we overlapped genes that both were downregulated and had an increased antisense RNA-seq signal over the gene bodies. Our analysis revealed that 14% of the downregulated genes also display increased antisense RNAs, suggesting a relatively spread interfering effect (Figure S2I). Overall, we conclude that activation of silent genes and interference is a very likely consequence of the YFFF mutations, but this point will require further investigation.

YFFF Transcription Yields Polyadenylated RNAs

An elevation of intergenic RNA levels at both gene ends could result not only from a transcriptional RT but also from an increased stabilization of the transcripts produced by natural RT (or both). To address this question, we purified chromatin-associated RNAs (chrRNAs) to perform chrRNA sequencing (chrRNA-seq) (Bhatt et al., 2012). This method allows scoring for nascent RNAs associated with chromatin and gives a similar readout compared to other nascent RNA-seq methods (Mayer et al., 2015; Nojima et al., 2015). Both individual and metagene profiling of this data, using the same gene body normalization approach as before, confirmed that the RT observed in the YFFF mutant originates at least from a transcriptional effect (Figures 3A–3C) both at the 3′ and 5′ ends. We also note that in both rWT and YFFF, chrRNAs tend to accumulate in 5′ antisense but not sense orientation.

Pol II ChIP-seq allowed further confirmation of the RT at 3′ ends with a delayed 3′ pause around 2.6 kb after the annotated 3′ ends (Figures 3D–3F). In these analyses, and as for RNA-seq, we rescaled the signals so that Pol II has comparable levels on the gene bodies (Figures 3D and S3A). At and after 3′ ends, we observed both an increased signal density (for at least 20 kb) and a delayed Pol II accumulation/pause occurring approximately 2.6 kb downstream of annotated 3′ ends. This delay is more pronounced and extends further than the one recently described for an Xrn2 D235A dominant mutation, also showing 3′ RT (Figure S3C) (Fong et al., 2015). We also note that the result remains clearly visible, even without performing the mentioned normalization, by just scaling the data to the same amount of sequenced tags (Figure S3B). We further calculated Pol II downstream (10 kb after annotated 3′ ends) RT indices and found significantly higher values in the mutant (Figure 3E, right panel), in contrast to little difference observed upstream of 5′ ends (Figure 3E, left panel). This latter result could be due to the fact that Pol II ChIP-seq is less sensitive over such large intervals to detect significant differences as compared to RNA-seq or chrRNA-seq.

Next, we wondered whether RT transcripts were polyadenylated, as the observed RT could arise from Pol II proceeding transcription following cleavage of the poly(A) transcripts, with no subsequent polyadenylation, as proposed in the torpedo termination model. To address this question, we performed poly(A) RNA-seq on rWT and YFFF cells and analyzed RT poly(A) transcription. As shown in Figures 3G and 3H and exemplified at the *CCR7* locus (Figure 3I), 3′ RT is clearly accompanied by apparent polyadenylation. Thus, the polyadenylation complex (CPA) might be associated with Pol II following the first poly(A) signal, generally located a bit before the end of 3′ UTRs. Intriguingly, increased poly(A) signal was also detected at 5′ ends of the genes in the antisense transcripts, indicating that the CPA could load at these locations. However, we cannot rule out that, despite two rounds of poly(A) RNA enrichment, the sequenced libraries may contain residual non-polyadenylated RNA. The observation that transcripts from non-polyadenylated histone genes show approximately 100-fold lower enrichment in our poly(A) RNA-seq data as compared to total RNA-seq data (Figure S3D) pleads, overall, against non-specific signal explaining our apparent RT phenotype. Currently, we cannot discriminate whether the high level of intergenic and antisense RT RNA in the poly(A) fraction of mutant YFFF originates from constant polyadenylation at cryptic poly(A) sites or from A-rich transcribed intergenic sequences. If polyadenylation of intergenic RNAs should occur, as our data suggest, this event obviously does not support termination of YFFF Pol II mutant. Together, our data indicate that both 5′ and 3′ RT in the YFFF are linked to a transcriptional effect and that these RNAs may be the subject of polyadenylation long after the normal poly(A) signal. We note, however, that polyadenylation can occur normally in the mutant despite the RT effect, as highlighted by the high density of reads at 3′ ends of genes, suggesting that polyadenylation and termination are uncoupled processes.

Tyrosine Mutations Are Associated with Reduced Pol II at the Promoter-Proximal Pause Sites and Reduced Nucleosome Depletion around Transcription Start Sites

Our Pol II ChIP-seq experiments showed a clear loss of Pol II accumulation in YFFF at promoters, as exemplified at the *MYCBP* locus (Figure 4A). This was also evidenced in metagene profile analyses by applying normalization to gene bodies (Figure 4B) as previously done. When indexing genes according to pausing score classes from low to high, we also found that reduced Pol II levels were more pronounced at highly paused genes in mutant YFFF (Figures 4C, 4D, and S4A). Such an effect was recently described following knockdown of the PAF1 complex, which also resulted in a global reduction of Pol II at pause sites in HCT116 cells (Chen et al., 2015). We then assessed whether this could be accompanied by a change in nucleosome occupancy at promoters and performed micrococcal nuclease (MNase) sequencing (MNase-seq) in both rWT and YFFF. Interestingly, we found that nucleosome densities in proximity of the nucleosome-depleted regions (NDRs), upstream of the transcription start sites (TSSs), were increased in the mutant (Figure 4E). This suggests that reduced Pol II levels at the pause site shorten the extent of NDRs and result in increased nucleosome occupancy, probably through reduced average Pol II occupancy. Analyzing our ChIP-seq and RNA-seq data, we asked whether reduced Pol II levels at promoters do correlate with the 3′ RT in the YFFF mutant (Figures S4B–S4D). This was not the case, as the 3 groups with low, medium, and high Pol II at the pause site showed similar effects. Our data support a global reduction of Pol II at promoter-proximal pause sites in the YFFF

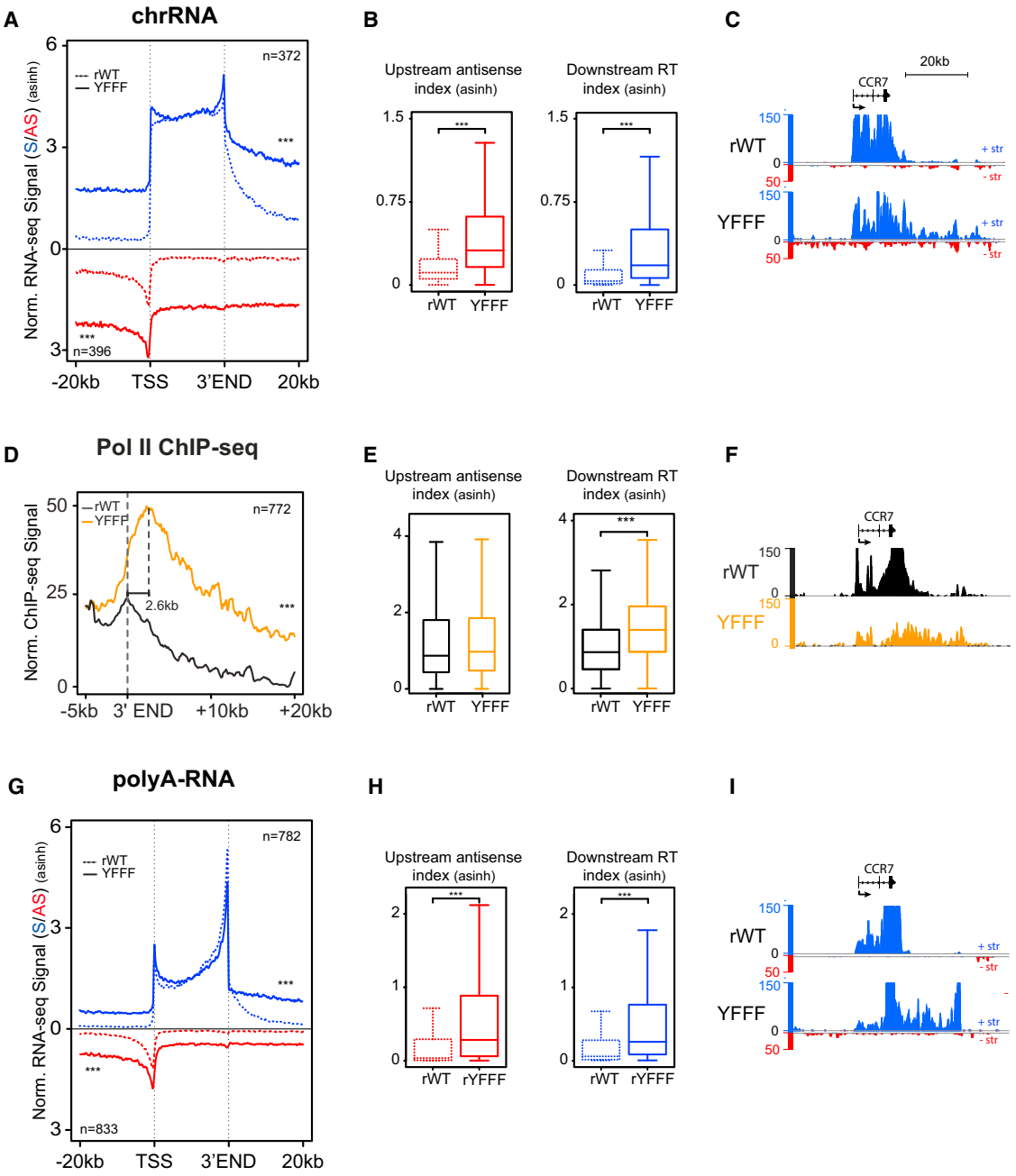


Figure 3. RT Phenotype of the YFFF Mutant Is due to Reduced Transcription Termination and Gives Rise to Polyadenylated Transcripts

(A) Average metagene profile of chrRNA-seq (asinh) in sense (blue) and antisense (red) orientation of the gene bodies and 20-kb upstream and downstream regions. Profiles were normalized so that sense RNA signals are equivalent on gene bodies.

(B) Boxplot of upstream antisense transcription index (left) and downstream RT transcription index (right) calculated with chrRNA signal.

(C) *CCR7* example of chrRNA-seq signal (y axis) RT in YFFF.

(D) Average Pol II ChIP-seq profiles of significantly bound genes in rWT (top 30% protein coding genes) in rWT and YFFF around the 3′ ends. Data are normalized to the same gene body level.

(E) Boxplot of upstream antisense transcription index (left) and downstream RT transcription index (right) calculated with Pol II ChIP signal.

(F) CCR7 locus showing Pol II RT activity (ChIP-seq signal is shown in the y axis).

(G) Average profile of sense and antisense poly(A) RNA signal in rWT and YFFF cells.

(H) Boxplot of upstream antisense transcription index (left) and downstream RT transcription index (right) calculated with poly(A) RNA.

(I) CCR7 locus showing that RT RNA is polyadenylated (RNA-seq signal in y axis).

***p values are < 2.2e−16. Figure S3E shows the non-normalized metagene chrRNA-seq and poly-RNA-seq show in (A) and (G). See also Figure S3.

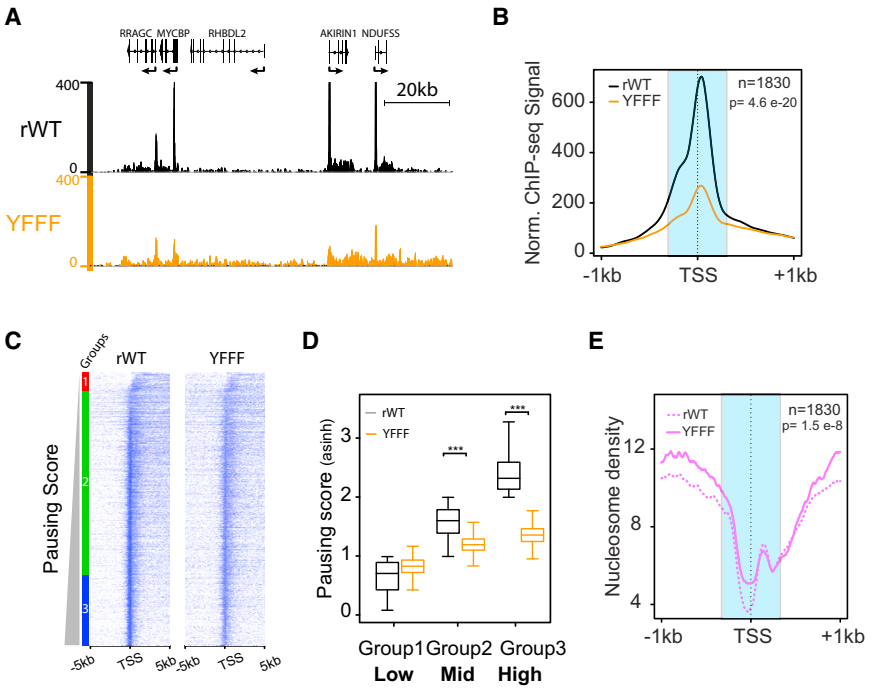


Figure 4. YFFF Mutations Result in Massive Loss of Pol II Accumulation at Promoter-Proximal Pause Sites

(A) Examples of Pol II ChIP-seq showing promoter-proximal pausing loss in YFFF cells occurring at multiple genes (ChIP-seq signal is shown in y axis).

(B) Average Pol II profiles at the TSSs of significantly bound genes (top 30% coding genes) in rWT and corresponding profile in YFFF. Data are normalized to bring the signals to the same level in gene body.

(C) Pol II density heatmaps at the TSSs of genes ranked by increasing pausing score in rWT and shown at rWT and YFFF TSSs. The boundaries of the 3 pausing groups 1–3 are shown on the left of the heatmaps.

(D) Boxplots of pausing scores for the 3 groups in rWT and YFFF. Only groups 2 and 3 show significant differences. Three asterisks indicate significantly different distributions with p value $< 2 \times 10^{-16}$.

(E) Nucleosome densities at promoters in rWT and the YFFF mutant. Data are normalized so that MNase-seq counts are equivalent in both experiments (scaling). In (B) and (E), the light blue rectangles indicate the areas that were taken in account for calculation of the indicated p values. See also Figure S4.

mutant. This reduced Pol II accumulation could equally result from a defect in initiation, pausing, or early elongation of the enzyme.

Transcribed Enhancers and Their Epigenetic Profiles Are Affected in the Mutant

We previously showed that Tyr1P of Pol II is enriched at transcribed enhancers (TEs) (Descostes et al., 2014). Importantly, TEs are more active and more tissue specific, but the act of transcription itself at these regions does not necessarily yield stable elongated RNAs (Lubas et al., 2015; Natoli and Andrau, 2012). Pausing of Pol II is also a hallmark of TEs (Core et al., 2014). The question, therefore, arises as to whether TEs also show reduced Pol II levels in the YFFF mutant, as is observed at promoters. To investigate this, we compared Pol II and epigenetic marks characteristic of regulatory regions at both promoters and TEs. We first isolated 1,316 intergenic TEs based on H3K27ac/H3K4me1/Pol II selection as described previously (Descostes et al., 2014), which we compared to a selection of active control promoters. Interestingly, enhancers showed a strong reduction of Pol II and a more modest but significant loss of H3K27ac and H3K4me1 when compared to promoters (see Figures 5A and S5A for examples). This effect was confirmed genome-wide at most isolated enhancers (Figures 5B and 5C) but did not hold true for H3K4me3, which remained comparable in rWT and YFFF. We did not observe significant alteration of nucleosome positioning or NDRs at enhancers. We also analyzed chrRNA-seq in this context and found little difference in rWT or YFFF, suggesting that the defect in Pol II density, most likely reflecting initiating or reduced Pol II pausing (Core et al., 2014), does not impair transcription at TEs (Figure S5B). This was in contrast with the situation observed at promoters, in which reduced pausing was detectable after RNA-seq and ChIP-seq profiling at the 5′ ends of the genes, in both the sense and antisense directions (Figures 3 and 4).

Together, the data indicate that promoters and enhancers display similar molecular phenotypes due to YFFF mutations: both show reduced accumulation of Pol II at the proximity of pause site.

Mutant YFFF Is Impaired in Its Interaction with the Med and Int Complexes

Our experiments shed light on a strong transcription termination defect phenotype at both the 5′ and 3′ ends. To get further insight into what could be the mechanism of tyrosine involvement in termination, we immunoprecipitated Pol II and analyzed its associated proteins in rWT and YFFF cells by mass spectrometry (MS) experiments. To improve the signal-to-noise ratio, we performed 5 biological replicates for each of the two pull-downs following induction of recombinant Rpb1 and inhibition of endogenous Pol II by α-amanitin. The results highlight a marked loss of the Med and Int complexes (Figure 6A), two major interactors of the Pol II CTD (Baillat et al., 2005; Conaway and Conaway, 2015). Most of the 31 and 12 subunits of the Med and Int complexes, respectively, were lost in the YFFF mutant and in all biological replicates (Tables S1 and S2). We note that subunits of the kinase module of Med were not associated with Pol II in rWT and YFFF cells. This is consistent with the observation that binding of CTD and the kinase module to Med is mutually exclusive (Allen and Taatjes, 2015; Tsai et al., 2013).

Because the cleavage/polyadenylation (CPA) complexes were previously linked to impaired termination phenotypes (Nojima et al., 2015), we searched for proteins associated with these complexes. We found that all of the CPSF subunits associated

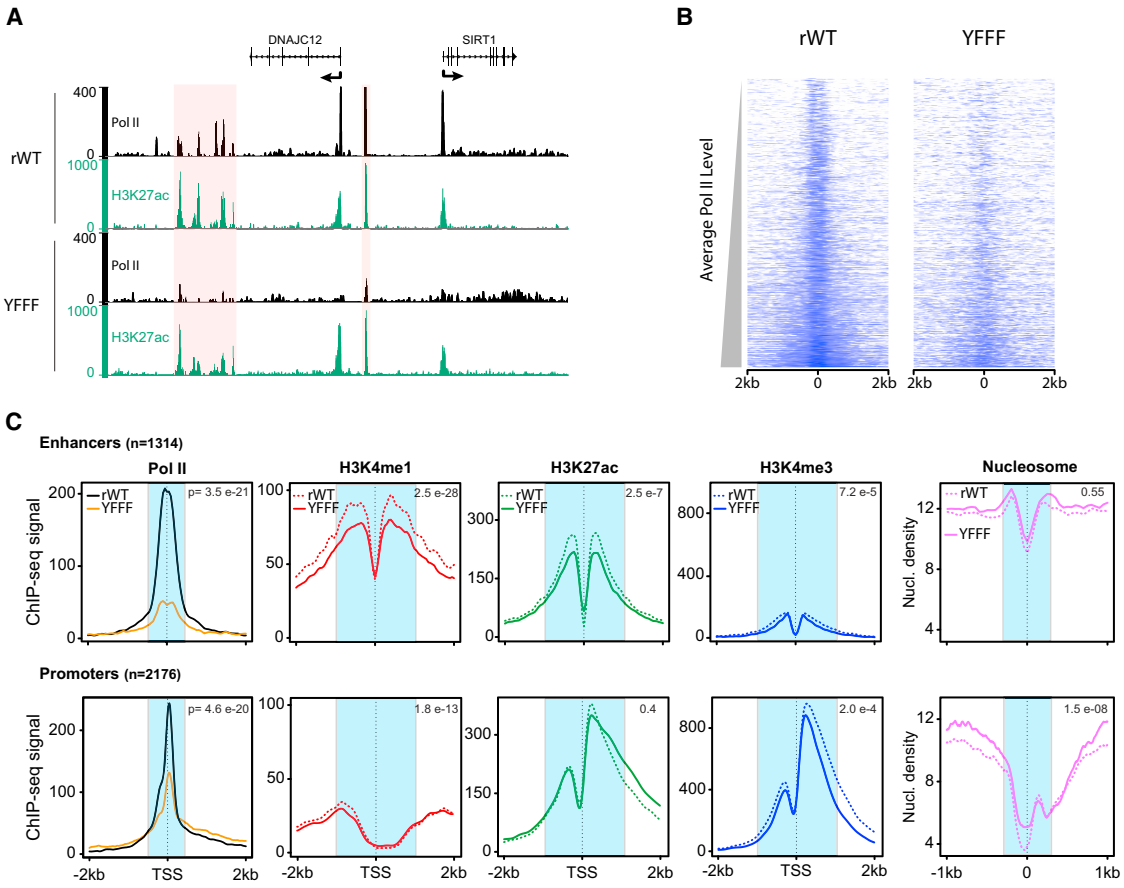


Figure 5. YFFF Mutations Result in Impaired Pol II Recruitment and Epigenetic Marking at Active Enhancers

(A) Putative enhancers or enhancer stretches (in pink rectangles) around the *DNAJC12* locus show altered Pol II loading in YFFF.

(B) Heatmap of Pol II densities at enhancers ranked by increasing Pol II signal in rWT and corresponding heatmap in YFFF.

(C) Average Pol II profiles, histone marks, and nucleosome density at intergenic enhancers (top row) and control promoters (bottom row; top 30% Pol II promoters in rWT without any other genes in surrounding 5-kb interval). The p values are indicated on the top right. Light blue rectangles indicate the areas that were taken in account for their calculation.

See also Figure S5.

with Pol II in rWT in MS were also associated with Pol II in the YFFF mutant. Furthermore, XRN2, as well as most splicing factors peptides, was found in comparable amounts in both fractions (Table S3).

We then asked whether the loss of interaction with Med and Int complexes was specific to mutation of the tyrosine residues and, again, made use of our S2AAA in which the same 3/4 of repeats are mutated as compared to YFFF. We analyzed the S2AAA mutant in MS experiments using the same induction/expression set-up. Interestingly, our results indicate that, while Int complex subunits remain associated with Pol II, many Med subunits are lost or show decreased interaction in the S2AAA mutant (Figure S6A; Tables S5 and S6), suggesting that the loss of interaction with Int might be more critical for the observed RT phenotype when tyrosine residues are mutated in the CTD.

To exclude the possibility that reduced Med and Int levels in MS were due to reduced protein expression of these complexes, we determined expression of the Med15 and Int11 subunits by western blot and found no significant difference (Figures S6B and S6C).

Among other interesting proteins that lost interaction with the YFFF Pol II mutant, we found two CTD phosphatase RPAP2-associated proteins, RPRD1a and RPRD2, and one subunit of the PAF1c (WDR61) complex (Table S2).

We next wondered whether the loss of association of Med and Int with Pol II in YFFF would result in their impaired recruitment on DNA. To address this question, we performed ChIP experiments in rWT and YFFF cells at several target characteristic locations (including promoters, enhancer, and snRNA gene). Following ChIP with Med1 and Ints11 antibodies (Abs), we found a decreased signal to background levels in YFFF, as compared to rWT cells, showing that lost Pol II contacts resulted in reduced Med/Int occupancy on DNA as well (Figure 6B).

Altogether, these data indicate that loss of interaction with Med and Int complexes is a major consequence of tyrosine mutations and strongly suggest that loss of one, the other, or both complexes might relate to the phenotypes linked to termination

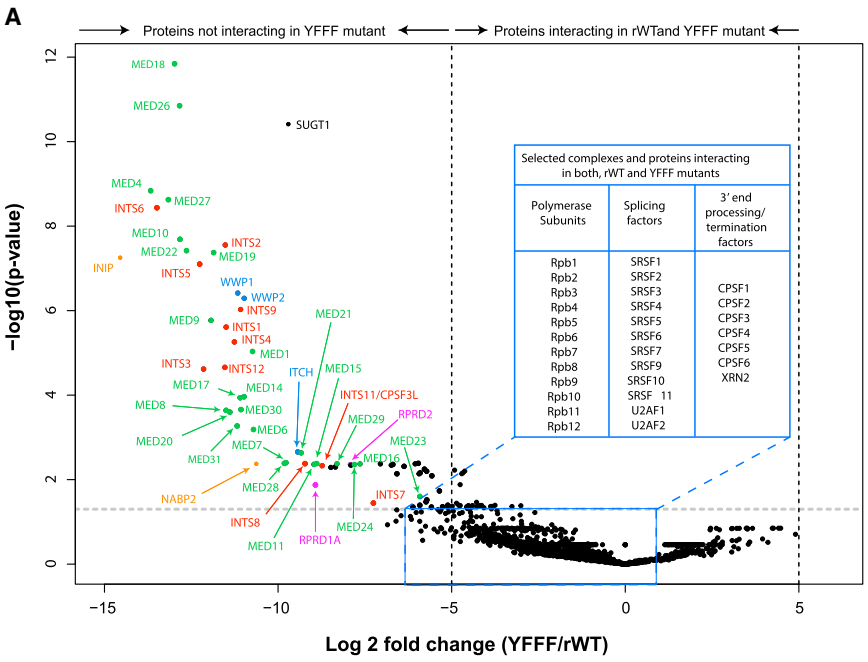


Figure 6. Mass Spectrometry Differential Analysis of rWT and YFFF Pol II Interactome

(A) Volcano plot comparing the Pol II interactome in rWT and the YFFF mutant. The table on the right lists selected proteins and complexes that interact with the Pol II of both rWT and the YFFF mutant. All 12 subunits of the Pol II, several splicing factors, and 3′ end processing factors are listed in the table. Represented on the left are the proteins and complexes that do not interact with the YFFF Pol II. Highlighted are the 25 subunits of the Med complex (green); 11 subunits of the Int (red), CTD phosphatase-associated proteins (magenta), and E3-ubiquitin ligase; and components of the SOSS complex and a few other proteins (blue). Threshold: $\log_2$ fold change ≥ 5; $p < 0.05$. Data are based on five independent biological replicates. For these experiments, cells were collected after 24 hr of induction and 48 hr of amanitin treatment as for the other assays.

(B) qPCR ChIP of Med (Med1) and Int (Inst11) at the *Ets1* enhancer (~24 kb upsteam) and at the *Rnu11, Myc, Snhg3-Rcc1, Kxd1*, and *Taf12* promoters in rWT and YFFF cells. Dashed lines highlight signals observed at the negative control region. Data are means ± SEM; n = 2.

See also Figure S6 and Tables S1, S2, and S3.

failure and promoter/enhancer defects. However, they might also relate to a yet-uncharacterized independent function of Tyr1.

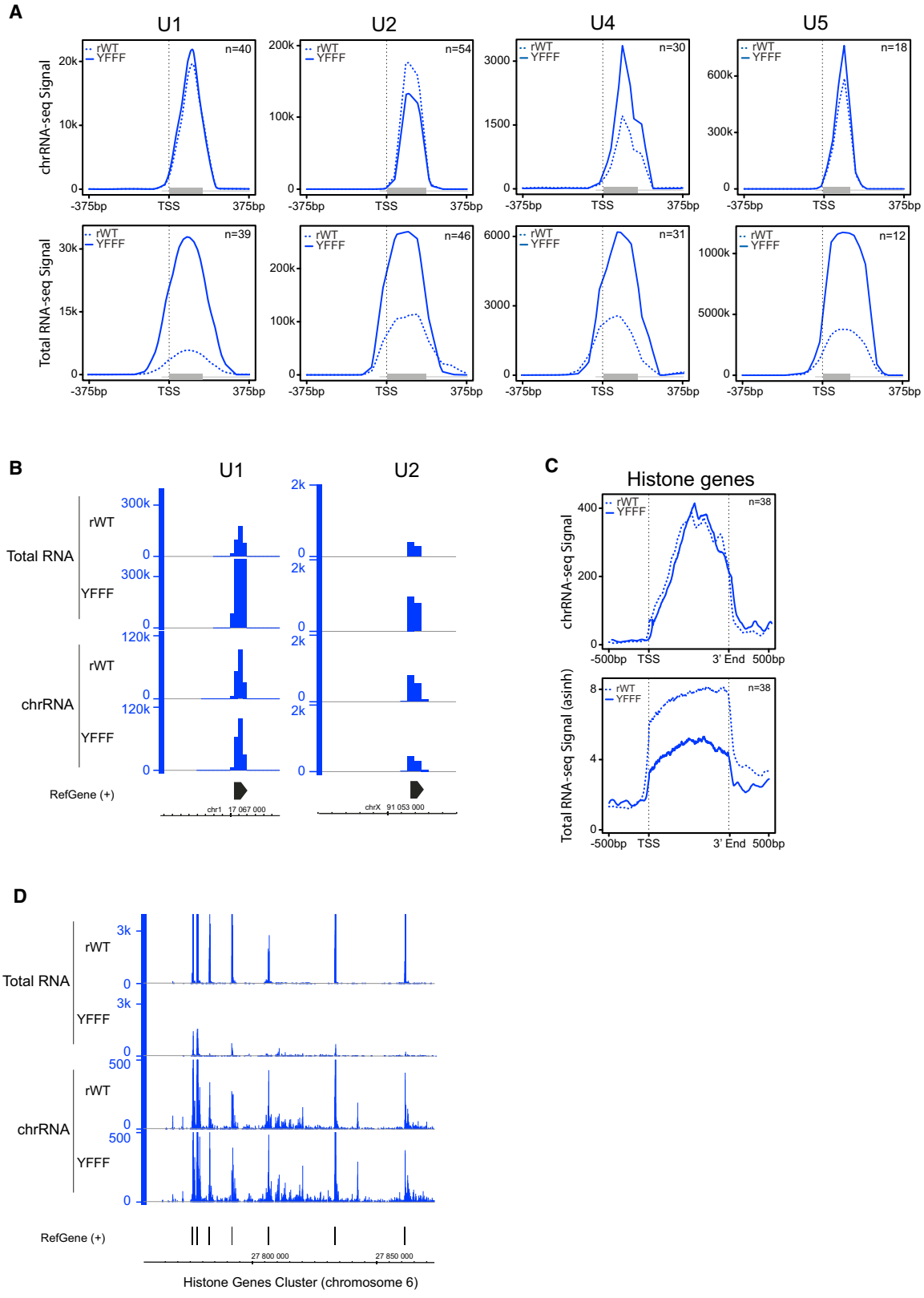
The YFFF Mutations Impair Maturation of snRNAs and Histone Non-polyadenylated Transcripts

The Int complex was previously described to function in synthesis and/or maturation of the snRNAs (Baillat et al., 2005; Egloff et al., 2007). Given the major interaction defect observed in our MS experiments, we wondered whether the YFFF mutant displayed impaired transcription/maturation at U1–U5 snRNA genes. We thus analyzed transcription at these genes. We found little effect for the YFFF mutant in chrRNA-seq experiments for U1, U2, and U5 genes and a slight RNA signal increase at the U4 genes (Figure 7A), suggesting that nascent transcription was, essentially, not affected for these genes. A rather important increase, varying from 3×–6×, was detected for total RNAs (Figure 7A), indicating that YFFF mutations may result in processing defects and stabilization of snRNAs. The opposite effect was observed for RNAs of non-poly(A) histone genes (Figure 7C). As for snRNA genes, nascent transcription (chrRNA-seq) was unaffected, but the total RNA-seq signal was reduced several-fold for histone RNAs (Figures 7C and 7D), indicating that non-poly(A) histone genes undergo massive destabilization and that proper processing of transcripts from histone genes is also affected in YFFF.

Overall, our results show strong and opposite stabilization/maturation defects for histone and snRNA genes in the YFFF mutant that might be associated with the loss of Int interaction. We conclude that the lack of tyrosine residues in the CTD can lead to the failure of specific CTD-coupled processes such as proper termination and processing of non-polyadenylated RNAs, while other processes such as transcript elongation or 3′ processing and polyadenylation of mRNAs remain unaffected.

DISCUSSION

In this article, we report a novel function of mammalian CTD for transcription termination at the 5′ and 3′ ends of genes. We show that Tyr1 residues of the CTD are required for termination, thereby strongly limiting the extent of pervasive transcription. Among the phenotypes of Tyr1 mutants analyzed in this study, the RT defect of the YFFF mutant was most striking. Transcription in this mutant remains high up to several hundreds of kilobases downstream of poly(A) sites, thus representing an exceptional case in which Pol II has lost the ability to terminate



(legend on next page)

transcription. Although we cannot completely rule out that Pol II association with a termination factor could be impaired due to its altered expression, we do not favor this possibility based on our MS data, where the number of CPSF peptides is comparable in both rWT and YFFF cells. Furthermore, in comparison to YFFF, the mutant S2AAA, in which Ser2 residues are mutated in the same distal CTD repeats, did not show significant RT, indicating the high specificity of our Tyr1 mutant phenotype. The observed RT phenotype of 5′ antisense transcripts in the YFFF mutant is consistent with our previous analysis showing the association of Tyr1P with antisense divergent transcription at the TSS of the genes (Descostes et al., 2014).

Pervasive transcription occurring after the 3′ ends of genes has been reported in specific WT cells or specific cellular context for Pol II mutants before. Recent works show that 3′ RT can be induced by osmotic stress or following herpes simplex 1 virus infection (Rutkowski et al., 2015; Vilborg et al., 2015). Given the overlap with the YFFF phenotype, it seems plausible that the CTD is involved in these processes by triggering transcriptional response to stress or viral infection. A mutation in the largest Pol II subunit was also described, resulting in a faster and less processive enzyme (Kaplan et al., 2008; Kireeva et al., 2008). This mutation provokes distal termination at many genes, which was correlated with the fast elongation rate of the mutant enzyme (Fong et al., 2015). Although this phenotype was less pronounced than the one described here, we do not exclude that Tyr mutations might also alter the velocity of Pol II. Our study also points out little, if any, termination defect in the S2AAA mutant. This is surprising, given a previous report that highlights the role of Ser2 for termination (Gu et al., 2013). However, we cannot rule out that the intact heptads (1–13) that were not mutated in our study still allow for the Ser2 to display a possible termination-related function.

How Pol II terminates transcription at the 3′ ends of genes remains a completely open question. Two main models were advanced in the past, the allosteric and the torpedo, the latter being prevalent in recent literature (Proudfoot, 2016). In this model, the exonuclease XRN2 attacks the uncapped 5′ end of the nascent RNA after 3′ cleavage and causes termination of transcription. Inactivation of XRN2 can result in termination defects downstream of poly(A) site and shift termination to further downstream sequences (Fong et al., 2015; West et al., 2004). However, XRN2 knockdown does not result in massive, genome-wide, pervasive transcription of intergenic sequences (Nojima et al., 2015), suggesting that XRN2 contributes to the tuning of termination but not to the removal of Pol II from the template. We also found that XRN2 recruitment to Pol II is not altered in the YFFF mutant (Figure 6), further supporting the notion that its association with Pol II cannot prevent pervasive transcription. More expectedly, knockdown of the CPA subunits CPSF73 (CPSF3) and CstF64 was shown to lead to reduced termination (Nojima et al., 2015) but did not result in a massive, pervasive transcription phenotype. Since the major CPSF subunits, including CPSF3, are recruited to Pol II, and since polyadenylation of RNA occurs, at least, to the same extent in YFFF mutant as compared to WT at the 3′ ends, we assume that the failure of termination occurs downstream of a functional 3′ processing machinery. Overall, we conclude that one of the main functions of the missing tyrosine residues in the CTD of the YFFF mutant is the control of transcription termination.

Another striking characteristic of the YFFF mutant is the reduced Pol II accumulation at the 5′ ends of genes. This could result from a loading defect of essential CTD-associated factors such as the ones we identified in our MS analyses, and we propose that an impaired promoter-proximal pausing could be the cause of the termination defect at the 5′ ends of the genes. We also found a delayed, but not decreased in amplitude, Pol II accumulation at the 3′ ends of genes, indicating that the lack of tyrosine residues generally affected pausing. At the 3′ ends, our result suggests that impaired complex(es) association with Pol II would not allow proper pausing of the enzyme at the first encountered poly(A) sites but, instead, at regions located, on average, 2.6 kb downstream. The consequence of this late pause could result in inefficient Pol II release from the template, possibly because of an impaired conformational transition in the enzyme or the lack of required signal, such as Tyr1P, required for efficient Pol II release from DNA. In either way, further studies will be required to address these possibilities.

Our MS experiments indicate that Int and Med complexes no longer associate with the CTD in the YFFF mutant. Both complexes were previously described as major CTD interactors, based on affinity purification (Baillat et al., 2005; Kim et al., 1994). The Med can act positively and negatively in the regulation of gene expression. It first supports the recruitment of Pol II to the promoter and later controls promoter release of Pol II in a CTD-dependent manner (Allen and Taatjes, 2015). This negative regulation of gene expression by Med was first described for mutants with truncated versions of CTD in yeast, which were able to maintain cell growth if specific subunits of the Med were mutated (Kim et al., 1994; Koleske and Young, 1994). Therefore, it appears likely that a potential promoter release phenotype observed in the YFFF mutant may be the consequence of the lack of interaction of CTD with the Med. This reduced pause could explain, at least in part, the 5′ antisense pervasive effect in the mutant. Studies in yeast and plants also proposed involvement of the MED18 subunit of the Med head module in transcription termination (Lai et al., 2014; Mukundan and Ansari, 2011). Finally, reports have described a possible role for Med and Int in Pol II release through recruitment of the super elongation complex (Donner et al., 2010; Gardini et al., 2014; Takahashi et al., 2011) and for Int in transcription termination (Skaar et al., 2015). Therefore, both Int and Med complexes

Figure 7. YFFF Mutations Affect Maturation of Transcripts from snRNA and Histone Genes

(A) Average metagene profile of chrRNA and total RNA (sense orientation) at the 50% most highly transcribed U1, U2, U4, and U5 snRNA genes. Gray rectangles indicate the corresponding gene size.

(B) Examples of total RNA and chrRNA signals at representative U1 and U2 snRNA loci.

(C) Average metagene profile of chrRNA and total RNA (asinh) in sense orientation at the 50% most highly transcribed non-polyadenylated histone genes.

(D) Example of total RNA and chrRNA signals at histone genes cluster.

were previously connected to transcription pausing- and termination-associated functions, making a direct link with the phenotypes described in this article.

At enhancers, we observed no apparent RT phenotype, unlike at 5′ or 3′ ends of genes. However, Pol II occupancy and acetylation of histone H3K27 were impaired, suggesting that pausing is affected at enhancers in mutant YFFF. Given the known similarities and differences of promoters and enhancers (Core et al., 2014; Koch et al., 2011), Tyr1 residues of CTD might provide regulatory information that has different consequences at promoters and enhancers. A recent report indicated that WDR82 knockdown in macrophages results in pervasive transcription at enhancers (Austenaa et al., 2015) that we did not observe at enhancers in the YFFF mutant. Conversely, when analyzing the transcriptome and Pol II data after WDR82 knockdown, we did not find strong RT at gene units (unpublished data). Altogether, this suggests that the control of termination might differ mechanistically for Pol II transcription initiated at promoters and enhancers.

Our work provides novel insights in the process of transcription termination and directly supports the involvement of Pol II CTD in this process. Future experiments should help to further dissect the mechanism of termination and establish possible roles of Med and Int complexes in termination and pause release. The YFFF mutant described here should also provide a great resource material to investigate the influence of extensive pervasive transcription on the frequency of DNA breaks in the genome, including the occurrence of DNA versus RNA polymerase collisions.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- CONTACT FOR REAGENT AND RESOURCE SHARING
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Establishing stable Cell lines
- METHOD DETAILS
  - I- Experimental Procedures
  - II- Bioinformatic Procedures
- QUANTIFICATION AND STATISTICAL ANALYSIS
- DATA AND SOFTWARE AVAILABILITY

SUPPLEMENTAL INFORMATION

Supplemental Information includes six figures and eight tables and can be found with this article online at https://doi.org/10.1016/j.molcel.2017.12.009.

ACKNOWLEDGMENTS

In the J.-C.A. lab, this work was supported by institutional grants from the CNRS and by specific grants from the "Agence Nationale de la Recherche" (ANR) and "Amorçage de Jeunes Équipes" Fondation pour la Recherche Médicale (FRM) AJE20130728183. In both the J.-C.A. and D.E. labs, the work was also supported by a German-French BMBF-ANR grant "EpiGlyco." This work was also supported by the Deutsche Forschungsgemeinschaft, SFB 1064 collaborative research center (CRC) – Chromatin Dynamics (to D.E.). We thank Eric Soler and Mounia Lagha for critical reading of this manuscript.

AUTHOR CONTRIBUTIONS

J.-C.A. and D.E. conceived the study and most of the experimental frame. N.S. made all constructs and performed all phenotypic characterization of the CTD mutants. M.A.M. and Y.Y. prepared chromatin extracts for ChIP and RNA (chrRNA) or total RNAs. M.A.M. and Y.Y. performed ChIP-seq and RNA-seq experiments, including QCs and library preparations. M.A.M., D.M., and A.Z.E.A. performed bioinformatics analyses. N.S., I.F., and A.I. performed MS experiments and data analysis. C.E. performed Med and Int ChIP experiments. T.-M.D. and R.S. contributed in the constructions and phenotypic characterization of the mutants. S.K. and H.B. performed the sequencing of the libraries. M.A.M., N.S., J.-C.A., and D.E. prepared and finalized the figures. J.-C.A. and D.E. wrote the manuscript, which was reviewed by all authors.

DECLARATION OF INTERESTS

The authors declare no competing financial interests.

Received: April 28, 2017
Revised: October 30, 2017
Accepted: December 11, 2017
Published: January 4, 2018

REFERENCES

Allen, B.L., and Taatjes, D.J. (2015). The Mediator complex: a central integrator of transcription. Nat. Rev. Mol. Cell Biol. *16*, 155–166.

Almada, A.E., Wu, X., Kriz, A.J., Burge, C.B., and Sharp, P.A. (2013). Promoter directionality is controlled by U1 snRNP and polyadenylation signals. Nature *499*, 360–363.

Anders, S., and Huber, W. (2010). Differential expression analysis for sequence count data. Genome Biol. *11*, R106.

Anders, S., Pyl, P.T., and Huber, W. (2015). HTSeq–a Python framework to work with high-throughput sequencing data. Bioinformatics *31*, 166–169.

Austenaa, L.M., Barozzi, I., Simonatto, M., Masella, S., Della Chiara, G., Ghisletti, S., Curina, A., de Wit, E., Bouwman, B.A., de Pretis, S., et al. (2015). Transcription of mammalian *cis*-regulatory elements is restrained by actively enforced early termination. Mol. Cell *60*, 460–474.

Baillat, D., Hakimi, M.A., Näär, A.M., Shilatifard, A., Cooch, N., and Shiekhattar, R. (2005). Integrator, a multiprotein mediator of small nuclear RNA processing, associates with the C-terminal repeat of RNA polymerase II. Cell *123*, 265–276.

Bartolomei, M.S., Halden, N.F., Cullen, C.R., and Corden, J.L. (1988). Genetic analysis of the repetitive carboxyl-terminal domain of the largest subunit of mouse RNA polymerase II. Mol. Cell. Biol. *8*, 330–339.

Bhatt, D.M., Pandya-Jones, A., Tong, A.J., Barozzi, I., Lissner, M.M., Natoli, G., Black, D.L., and Smale, S.T. (2012). Transcript dynamics of proinflammatory genes revealed by sequence analysis of subcellular RNA fractions. Cell *150*, 279–290.

Chapman, R.D., Heidemann, M., Albert, T.K., Mailhammer, R., Flatley, A., Meisterernst, M., Kremmer, E., and Eick, D. (2007). Transcribing RNA polymerase II is phosphorylated at CTD residue serine-7. Science *318*, 1780–1782.

Chen, F.X., Woodfin, A.R., Gardini, A., Rickels, R.A., Marshall, S.A., Smith, E.R., Shiekhattar, R., and Shilatifard, A. (2015). PAF1, a molecular regulator of promoter-proximal pausing by RNA polymerase II. Cell *162*, 1003–1015.

Conaway, R.C., and Conaway, J.W. (2015). Orchestrating transcription with the Pol II CTD. Nat. Rev. Mol. Cell Biol. *16*, 128.

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

Core, L.J., Martins, A.L., Danko, C.G., Waters, C.T., Siepel, A., and Lis, J.T. (2014). Analysis of nascent RNA identifies a unified architecture of initiation regions at mammalian promoters and enhancers. Nat. Genet. *46*, 1311–1320.

Descostes, N., Heidemann, M., Spinelli, L., Schüller, R., Maqbool, M.A., Fenouil, R., Koch, F., Innocenti, C., Gut, M., Gut, I., et al. (2014). Tyrosine

phosphorylation of RNA polymerase II CTD is associated with antisense promoter transcription and active enhancers in mammalian cells. eLife *3*, e02105.

Donner, A.J., Ebmeier, C.C., Taatjes, D.J., and Espinosa, J.M. (2010). CDK8 is a positive regulator of transcriptional elongation within the serum response network. Nat. Struct. Mol. Biol. *17*, 194–201.

Egloff, S., O'Reilly, D., Chapman, R.D., Taylor, A., Tanzhaus, K., Pitts, L., Eick, D., and Murphy, S. (2007). Serine-7 of the RNA polymerase II CTD is specifically required for snRNA gene expression. Science *318*, 1777–1779.

Eick, D., and Geyer, M. (2013). The RNA polymerase II carboxy-terminal domain (CTD) code. Chem. Rev. *113*, 8456–8490.

Fenouil, R., Cauchy, P., Koch, F., Descostes, N., Cabeza, J.Z., Innocenti, C., Ferrier, P., Spicuglia, S., Gut, M., Gut, I., and Andrau, J.C. (2012). CpG islands and GC content dictate nucleosome depletion in a transcription-independent manner at mammalian promoters. Genome Res. *22*, 2399–2408.

Fenouil, R., Descostes, N., Spinelli, L., Koch, F., Maqbool, M.A., Benoukraf, T., Cauchy, P., Innocenti, C., Ferrier, P., and Andrau, J.-C. (2016). Pasha: a versatile R package for piling chromatin HTS data. Bioinformatics *32*, 2528–2530.

Fong, N., Brannan, K., Erickson, B., Kim, H., Cortazar, M.A., Sheridan, R.M., Nguyen, T., Karp, S., and Bentley, D.L. (2015). Effects of transcription elongation rate and Xrn2 exonuclease activity on RNA polymerase II termination suggest widespread kinetic competition. Mol. Cell *60*, 256–267.

Gardini, A., Baillat, D., Cesaroni, M., Hu, D., Marinis, J.M., Wagner, E.J., Lazar, M.A., Shilatifard, A., and Shiekhattar, R. (2014). Integrator regulates transcriptional initiation and pause release following activation. Mol. Cell *56*, 128–139.

Grosso, A.R., Leite, A.P., Carvalho, S., Matos, M.R., Martins, F.B., Vítor, A.C., Desterro, J.M., Carmo-Fonseca, M., and de Almeida, S.F. (2015). Pervasive transcription read-through promotes aberrant expression of oncogenes and RNA chimeras in renal carcinoma. eLife *4*, e09214.

Gu, B., Eick, D., and Bensaude, O. (2013). CTD serine-2 plays a critical role in splicing and termination factor recruitment to RNA polymerase II in vivo. Nucleic Acids Res. *41*, 1591–1603.

Harlen, K.M., Trotta, K.L., Smith, E.E., Mosaheb, M.M., Fuchs, S.M., and Churchman, L.S. (2016). Comprehensive RNA polymerase II interactomes reveal distinct and varied roles for each phospho-CTD residue. Cell Rep. *15*, 2147–2158.

Hazelbaker, D.Z., Marquardt, S., Wlotzka, W., and Buratowski, S. (2013). Kinetic competition between RNA polymerase II and Sen1-dependent transcription termination. Mol. Cell *49*, 55–66.

Hintermair, C., Heidemann, M., Koch, F., Descostes, N., Gut, M., Gut, I., Fenouil, R., Ferrier, P., Flatley, A., Kremmer, E., et al. (2012). Threonine-4 of mammalian RNA polymerase II CTD is targeted by Polo-like kinase 3 and required for transcriptional elongation. EMBO J. *31*, 2784–2797.

Hsin, J.P., Sheth, A., and Manley, J.L. (2011). RNAP II CTD phosphorylated on threonine-4 is required for histone mRNA 3′ end processing. Science *334*, 683–686.

Hsin, J.P., Li, W., Hoque, M., Tian, B., and Manley, J.L. (2014). RNAP II CTD tyrosine 1 performs diverse functions in vertebrate cells. eLife *3*, e02112.

Iben, J.R., Mazeika, J.K., Hasson, S., Rijal, K., Arimbasseri, A.G., Russo, A.N., and Maraia, R.J. (2011). Point mutations in the Rpb9-homologous domain of Rpc11 that impair transcription termination by RNA polymerase III. Nucleic Acids Res. *39*, 6100–6113.

Ishihama, Y., Rappsilber, J., and Mann, M. (2006). Modular stop and go extraction tips with stacked disks for parallel and multidimensional peptide fractionation in proteomics. J. Proteome Res. *5*, 988–994.

Kaplan, C.D., Larsson, K.M., and Kornberg, R.D. (2008). The RNA polymerase II trigger loop functions in substrate selection and is directly targeted by alpha-amanitin. Mol. Cell *30*, 547–556.

Kim, Y.J., Björklund, S., Li, Y., Sayre, M.H., and Kornberg, R.D. (1994). A multiprotein mediator of transcriptional activation and its interaction with the C-terminal repeat domain of RNA polymerase II. Cell *77*, 599–608.

Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R., and Salzberg, S.L. (2013). TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biol. *14*, R36.

Kireeva, M.L., Nedialkov, Y.A., Cremona, G.H., Purtov, Y.A., Lubkowska, L., Malagon, F., Burton, Z.F., Strathern, J.N., and Kashlev, M. (2008). Transient reversal of RNA polymerase II active site closing controls fidelity of transcription elongation. Mol. Cell *30*, 557–566.

Koch, F., Fenouil, R., Gut, M., Cauchy, P., Albert, T.K., Zacarias-Cabeza, J., Spicuglia, S., de la Chapelle, A.L., Heidemann, M., Hintermair, C., et al. (2011). Transcription initiation platforms and GTF recruitment at tissue-specific enhancers and promoters. Nat. Struct. Mol. Biol. *18*, 956–963.

Koleske, A.J., and Young, R.A. (1994). An RNA polymerase II holoenzyme responsive to activators. Nature *368*, 466–469.

Lai, Z., Schluttenhofer, C.M., Bhide, K., Shreve, J., Thimmapuram, J., Lee, S.Y., Yun, D.J., and Mengiste, T. (2014). MED18 interaction with distinct transcription factors regulates multiple plant functions. Nat. Commun. *5*, 3064.

Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nat. Methods *9*, 357–359.

Lepoivre, C., Belhocine, M., Bergon, A., Griffon, A., Yammine, M., Vanhille, L., Zacarias-Cabeza, J., Garibal, M.A., Koch, F., Maqbool, M.A., et al. (2013). Divergent transcription is associated with promoters of transcriptional regulators. BMC Genomics *14*, 914.

Lubas, M., Andersen, P.R., Schein, A., Dziembowski, A., Kudla, G., and Jensen, T.H. (2015). The human nuclear exosome targeting complex is loaded onto newly synthesized RNA to direct early ribonucleolysis. Cell Rep. *10*, 178–192.

Mayer, A., Heidemann, M., Lidschreiber, M., Schreieck, A., Sun, M., Hintermair, C., Kremmer, E., Eick, D., and Cramer, P. (2012). CTD tyrosine phosphorylation impairs termination factor recruitment to RNA polymerase II. Science *336*, 1723–1725.

Mayer, A., di Iulio, J., Maleri, S., Eser, U., Vierstra, J., Reynolds, A., Sandstrom, R., Stamatoyannopoulos, J.A., and Churchman, L.S. (2015). Native elongating transcript sequencing reveals human transcriptional activity at nucleotide resolution. Cell *161*, 541–554.

McCracken, S., Fong, N., Yankulov, K., Ballantyne, S., Pan, G., Greenblatt, J., Patterson, S.D., Wickens, M., and Bentley, D.L. (1997). The C-terminal domain of RNA polymerase II couples mRNA processing to transcription. Nature *385*, 357–361.

Meininghaus, M., Chapman, R.D., Horndasch, M., and Eick, D. (2000). Conditional expression of RNA polymerase II in mammalian cells. Deletion of the carboxyl-terminal domain of the large subunit affects early steps in transcription. J. Biol. Chem. *275*, 24375–24382.

Mukundan, B., and Ansari, A. (2011). Novel role for mediator complex subunit Srb5/Med18 in termination of transcription. J. Biol. Chem. *286*, 37053–37057.

Nagaike, T., Logan, C., Hotta, I., Rozenblatt-Rosen, O., Meyerson, M., and Manley, J.L. (2011). Transcriptional activators enhance polyadenylation of mRNA precursors. Mol. Cell *41*, 409–418.

Natoli, G., and Andrau, J.C. (2012). Noncoding transcription at enhancers: general principles and functional models. Annu. Rev. Genet. *46*, 1–19.

Nojima, T., Gomes, T., Grosso, A.R.F., Kimura, H., Dye, M.J., Dhir, S., Carmo-Fonseca, M., and Proudfoot, N.J. (2015). Mammalian NET-seq reveals genome-wide nascent transcription coupled to RNA processing. Cell *161*, 526–540.

Ntini, E., Järvelin, A.I., Bornholdt, J., Chen, Y., Boyd, M., Jørgensen, M., Andersson, R., Hoof, I., Schein, A., Andersen, P.R., et al. (2013). Polyadenylation site-induced decay of upstream transcripts enforces promoter directionality. Nat. Struct. Mol. Biol. *20*, 923–928.

Pinto, P.A., Henriques, T., Freitas, M.O., Martins, T., Domingues, R.G., Wyrzykowska, P.S., Coelho, P.A., Carmo, A.M., Sunkel, C.E., Proudfoot, N.J., and Moreira, A. (2011). RNA polymerase II kinetics in polo polyadenylation signal selection. EMBO J. *30*, 2431–2444.

Preker, P., Nielsen, J., Kammler, S., Lykke-Andersen, S., Christensen, M.S., Mapendano, C.K., Schierup, M.H., and Jensen, T.H. (2008). RNA exosome depletion reveals transcription upstream of active human promoters. Science *322*, 1851–1854.

Proudfoot, N.J. (2016). Transcriptional termination in mammals: stopping the RNA polymerase II juggernaut. Science *352*, aad9926.

Rutkowski, A.J., Erhard, F., L'Hernault, A., Bonfert, T., Schilhabel, M., Crump, C., Rosenstiel, P., Efstathiou, S., Zimmer, R., Friedel, C.C., and Dölken, L. (2015). Widespread disruption of host transcription termination in HSV-1 infection. Nat. Commun. *6*, 7126.

Schlackow, M., Nojima, T., Gomes, T., Dhir, A., Carmo-Fonseca, M., and Proudfoot, N.J. (2017). Distinctive patterns of transcription and RNA processing for human lincRNAs. Mol. Cell *65*, 25–38.

Seila, A.C., Calabrese, J.M., Levine, S.S., Yeo, G.W., Rahl, P.B., Flynn, R.A., Young, R.A., and Sharp, P.A. (2008). Divergent transcription from active promoters. Science *322*, 1849–1851.

Shaaban, S.A., Krupp, B.M., and Hall, B.D. (1995). Termination-altering mutations in the second-largest subunit of yeast RNA polymerase III. Mol. Cell. Biol. *15*, 1467–1478.

Shevchenko, A., Chernushevich, I., Wilm, M., and Mann, M. (2000). De Novo peptide sequencing by nanoelectrospray tandem mass spectrometry using triple quadrupole and quadrupole/time-of-flight instruments. Methods Mol. Biol. *146*, 1–16.

Skaar, J.R., Ferris, A.L., Wu, X., Saraf, A., Khanna, K.K., Florens, L., Washburn, M.P., Hughes, S.H., and Pagano, M. (2015). The Integrator complex controls the termination of transcription at diverse classes of gene targets. Cell Res. *25*, 288–305.

Takahashi, H., Parmely, T.J., Sato, S., Tomomori-Sato, C., Banks, C.A., Kong, S.E., Szutorisz, H., Swanson, S.K., Martin-Brown, S., Washburn, M.P., et al. (2011). Human mediator subunit MED26 functions as a docking site for transcription elongation factors. Cell *146*, 92–104.

Tsai, K.L., Sato, S., Tomomori-Sato, C., Conaway, R.C., Conaway, J.W., and Asturias, F.J. (2013). A conserved Mediator-CDK8 kinase module association regulates Mediator-RNA polymerase II interaction. Nat. Struct. Mol. Biol. *20*, 611–619.

Vilborg, A., Passarelli, M.C., Yario, T.A., Tycowski, K.T., and Steitz, J.A. (2015). Widespread inducible transcription downstream of human genes. Mol. Cell *59*, 449–461.

West, S., Gromak, N., and Proudfoot, N.J. (2004). Human 5′ → 3′ exonuclease Xrn2 promotes transcription termination at co-transcriptional cleavage sites. Nature *432*, 522–525.

Wilm, M., Shevchenko, A., Houthaeve, T., Breit, S., Schweigerer, L., Fotsis, T., and Mann, M. (1996). Femtomole sequencing of proteins from polyacrylamide gels by nano-electrospray mass spectrometry. Nature *379*, 466–469.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Rabbit polyclonal anti-H3K4me1	Abcam	ab8895
Rabbit polyclonal anti-H3K4me3	Abcam	ab8580
Rabbit polyclonal anti-H3K27ac	Abcam	ab4729
Rabbit polyclonal anti-HA	Abcam	ab9110
Rat monoclonal anti-HA	Roche	3F10
Rabbit polyclonal anti-MED15	Proteintech	11566-1-AP
Rabbit polyclonal anti-Med1	Bethyly Labs	A301-793A
Rabbit polyclonal anti-INT11	Bethyl Labs	A301-274A
Mouse Monoclonal anti-RPB1	Elisabeth Kremmer, Helmholtz Zentrum, Munich	Pol3.3
GAPDH	Elisabeth Kremmer, Helmholtz Zentrum, Munich	5C4
Rat monoclonal anti-Ser2P	Helmholtz Zentrum Munich	3E10
Rat monoclonal anti-Ser5P	Helmholtz Zentrum Munich	3E8
Rat monoclonal anti-Ser7P	Helmholtz Zentrum Munich	4E12
Rat monoclonal anti-Thr4P	Helmholtz Zentrum Munich	6D7
Chemicals, Peptides, and Recombinant Proteins		
Micrococcal nuclease	Roche	10107921001
RNaseIII	Thermo Fisher	AM2290
Turbo DNA-Free	Thermo Fisher	AM1907
Alpha-amanitin	Sigma Aldrich	A2263
G-418 Solution	Sigma Aldrich	0000004727878001
Deposited Data		
ChIP-seq, RNA-seq and MNase-seq data	This study	GSE94330
Experimental Models: Cell Lines		
Raji cells	ATCC	CCL-86
Recombinant DNA		
Mouse Rpb1 gene cloned into LS*mock vector	Meininghaus et al.,2000	rWT
Mouse Rpb1 gene cloned into LS*mock vector	Meininghaus et al.,2000	YFFY
Mouse Rpb1 gene cloned into LS*mock vector	Meininghaus et al.,2000	YYFF
Mouse Rpb1 gene cloned into LS*mock vector	Meininghaus et al.,2000	FYYF
Mouse Rpb1 gene cloned into LS*mock vector	Meininghaus et al.,2000	YFFF
Mouse Rpb1 gene cloned into LS*mock vector	Meininghaus et al.,2000	S2AAA
Oligonucleotides		
SNHG3-prom-F: GTGGTCGCTTCTTCTCCTTG	This study	
SNHG3-prom-R: TAGGGAAGCTCGGCTACTGA	This study	
ETS1-Enh-UPS-1-F: GGCTGTTCGTCTCCCAAGTA	This study	
ETS1-Enh-UPS-1-R: CACTGCAGGTGGTAATTTGC	This study	
Myc-prom-F: AGGGATCGCGCTGAGTATAA	This study	
Myc-prom-R: TGCCTCTCGCTGGAATTACT	This study	
TAF12-prom-F: ACCTGGTCCTTCGAACACTG	This study	
TAF12-prom-R: GGCAGTTGAGGAACAAGAGC	This study	
Rnu11-prom-F: ACCCTGCTTTGGTGACAGAG	This study	

(*Continued on next page*)

Continued		
REAGENT or RESOURCE	SOURCE	IDENTIFIER
Rnu11-prom-R: ATCACCAGCTGCCCAAATAC	This study	
Kxd1-prom-F: CAAAAGTGGAGCAGGGATGT	This study	
Kxd1-prom-R: CCCCAAGGTCGTAAATGCTA	This study	
Software and Algorithms		
PASHA	Fenouil et al., 2016	https://cran.r-project.org/web/packages/Pasha/
Bowtie2	Langmead and Salzberg, 2012	http://bowtie-bio.sourceforge.net/bowtie2/index.shtml
TopHat2	Kim et al., 2013	http://ccb.jhu.edu/software

CONTACT FOR REAGENT AND RESOURCE SHARING

Further information and requests for reagents should be directed to and will be fulfilled by Lead Contact Jean-Christophe Andrau (jean-christophe.andrau@igmm.cnrs.fr).

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Establishing stable Cell lines

Raji is an Epstein-Barr-virus-positive Burkitt's lymphoma cell line of Male origin. Full-length Rpb1 expression vector (rWT, YFFF, YYFF, FYYF and YFFY) were transfected into Raji cells using 1 X 10^7 cells (10 μg plasmid, 960 μF, 250V). Polyclonal cell lines were established after selection with G418 (1 mg/ml) for 2-3 weeks. Tetracycline was removed to induce the expression of recombinant Rpb1 by washing the cells three times with 50 mL of phosphate-buffered saline (PBS) supplemented with 1% fetal calf serum (FCS) (GIBCO, Invitrogen). 24 h after induction, cells were cultured in the presence of 2 μg/ml of α-amanitin (Sigma) to inhibit endogenous Pol II. Cell lines were maintained in RPMI 1640 medium supplemented with 10% fetal calf serum, 100 U/ml penicillin, 100 μg/ml streptomycin, 2mM L-glutamine (GIBCO, Invitrogen) at 37°C and 5% CO2.

METHOD DETAILS

I- Experimental Procedures

Antibodies

Monoclonal antibodies specific for haemagglutinin (HA)-tag (3F10, Roche), (12CA5, Sigma) and polyclonal antibodies against MED15 (11566-1-AP, Proteintech), MED1 (A301-793A, Bethyl laboratories) and INT11 (A301-274A, Bethyl laboratories) are commercially available. Monoclonal antibodies against Rpb1 (Pol 3.3), Ser2P (3E10), Ser5P (3E8), Ser7P (4E12) and Thr4P (6D7) were described previously (Chapman et al., 2007; Hintermair et al., 2012) and monoclonal antibody against GAPDH (5C4) was received from Elisabeth Kremmer, Helmholtz Zentrum Munich.

Construction of the CTD mutants

Mouse CTD sequences of rWT and tyrosine mutants (YYYF, YYFF, FYYF and YFFY) with an optimized human codon usage were synthesized by Gene Art (Regensburg) and cloned into LS*mock vector (Meininghaus et al., 2000). All final constructs were sequenced before usage.

Western blot analysis

Cells were washed twice with PBS and directly lysed with 2X Laemmli buffer. Whole cell lysates were separated on SDS-PAGE (6.5% gel) and blotted on a nitrocellulose membrane (GE healthcare). The membranes were blocked with 5% milk/TBS-T solution for 1 h and incubated overnight with the primary antibody at 4°C. Afterward, the membranes were incubated, either with IRDye-labeled secondary antibodies against rat (680 nm; Alexa, Invitrogen) and/or mouse (800 nm; Rockford, Biomol) and analyzed using an Odyssey Imaging System (Li-Cor) or they were stained with HRP-conjugated secondary antibodies against rat (Sigma) or mouse (Promega) to be detected by chemiluminescence.

Growth kinetics

Growth kinetics of rWT, CTD mutants and wild-type Raji cells were monitored over a period of 10 days. For each cell line, 20 X 10^6 cells were induced and the number of living cells (Nl) and the number of dead cells (Nd) were calculated every day using trypan-blue staining. Cumulative living cell number was calculated by multiplying the total number of living cells (Nl) with the factor by which the culture was split over the course of the experiment. These kinetics recapitulate growth features of samples collection for ChIPseq and RNaseq experiments.

Purification of Pol II interacting proteins for mass spectrometric analysis

For purification of recombinant Rpb1, α-HA antibody (12CA5) was coupled to Sepharose A/G beads for 4 h at 4°C. Simultaneously, cells (7.5 X 10^7) were washed twice with ice cold PBS and lysed in lysis buffer [50mM Tris-HCl pH 8.0, 150mM NaCl, 1% NP-40 (Roche), 1 X PhosStop (Roche), 1 X protease cocktail (Roche)] for 30 min on ice. Samples were sonified (Sonifier 250 BRANSON, 3 × 20 cycles, output 5, duty cycle 50) and incubated on a shaker for 1 h at 4°C. Samples were then centrifuged at 10,000 g for 15 minutes and the supernatants were incubated with antibody-coupled Sepharose A/G beads for overnight at 4°C. Next day, beads were washed three times with lysis buffer and continued with either on-beads trypsin digest or boiled with 2X Laemmli buffer (95°C, 8 min) to load proteins on SDS-PAGE for the subsequent in-gel trypsin digest.

On-beads trypsin digest

Following the standard immunoprecipitation procedure, beads were first washed with lysis buffer (three times) and then with 50mM NH_4HCO_3 (ammonium bicarbonate). For trypsin digest, beads were incubated with 100 μl of 10 ng/μl of trypsin solution in 1M Urea and 50mM NH_4HCO_3 for 30 minutes at 25°C. The supernatant was collected, beads washed twice with 50mM NH_4HCO_3 and all three supernatants collected together and incubated overnight at 25°C after addition of 1mM DTT. 27mM of iodoacetamide (IAA) was then added to the samples and incubated at 25°C for 30 minutes in dark. Next, 1 μl of 1M DTT was added to the samples and incubated for 10 minutes to quench the IAA. Finally, 2.5 μl of trifluoroacetic acid (TFA) were added to the samples and desalted using C18 stage tips (Ishihama et al., 2006). Samples were evaporated to dryness, re-suspended in 30 μl of 0.1% formic acid solution and stored at −20°C until LC-MS analysis.

In-gel trypsin digest

A standardized protocol was used for in-gel digestion with minor modifications (Shevchenko et al., 2000; Wilm et al., 1996). The digested peptides were evaporated to 5 μl and re-suspended in 30 μl of 0.1% TFA solution prior to desalting by C18 stage tips. Samples were evaporated to dryness and re-suspended in 30 μl of 0.1% formic acid solution and stored at −20°C until LC-MS analysis.

Liquid Chromatography Coupled to Tandem Mass Spectrometry

For LC-MS/MS purposes, desalted peptides were injected in an Ultimate 3000 RSLCnano system (Thermo), separated in a 15-cm analytical column (75 μm ID with ReproSil-Pur C18-AQ 2.4 μm from Dr. Maisch) with a 50 min gradient from 5 to 60% acetonitrile in 0.1% formic acid. The effluent from the HPLC was directly electrosprayed into a QexactiveHF (Thermo) operated in data dependent mode to automatically switch between full scan MS and MS/MS acquisition. Survey full scan MS spectra (from m/z 375–1600) were acquired with resolution R = 60,000 at m/z 400 (AGC target of 3x106). The 10 most intense peptide ions with charge states between 2 and 5 were sequentially isolated to a target value of $1x10^5$, and fragmented at 27% normalized collision energy. Typical mass spectrometric conditions were: spray voltage, 1.5 kV; no sheath and auxiliary gas flow; heated capillary temperature, 250°C; ion selection threshold, 33.000 counts. MaxQuant 1.5.2.8 was used to identify proteins and quantify by iBAQ with the following parameters: Database, Uniprot_Hsapiens_3AUP000005640_151111; MS tol, 10ppm; MS/MS tol, 0.5 Da; Peptide FDR, 0.1; Protein FDR, 0.01 Min. peptide Length, 5; Variable modifications, Oxidation (M); Fixed modifications, Carbamidomethyl (C); Peptides for protein quantitation, razor and unique; Min. peptides, 1; Min. ratio count, 2. Identified proteins were considered as interaction partners if their MaxQuant iBAQ values displayed a greater than log2 5-fold enrichment and p value 0.05 (ANOVA) when compared to the rWT control. The data were processed for visualization using R (https://www.r-project.org/).

ChIP-seq and ChIP-qPCR

To cross-link the cells for ChIP, $1/10^{th}$ volume of 10X crosslinking solution (100mM NaCl, 1mM EDTA pH 8, 0.5mM EGTA pH 8, 50mM HEPES pH 7.8 and 11% formaldehyde) was added to the raji cells in culture medium. After 10 minutes' incubation at room temperature, glycine was added to a final concentration of 250mM to quench the remaining formaldehyde and stop cross-linking. After five minutes of quenching, cells were washed twice with cold PBS. Cells were then sonicated as described in next paragraph or snap frozen in liquid nitrogen and stored at −80°C for sonication at a later stage.

For sonication, 50 $x10^6$ cross-linked raji cells were lysed by resuspending in cold 2.5mL LB1 (50mM HEPES pH 7.5, 140mM NaCl, 1mM EDTA pH 8, 10% glycerol, 0.75% NP-40, 0.25% Triton X-100) at 4°C for 20 minutes on a rotating wheel. Nuclei were pelleted down by spinning at 1350 *rcf.* in a refrigerated centrifuge and washed in 2.5mL LB2 (200mM NaCl, 1mM EDTA pH 8, 0.5mM EGTA pH 8, 10mM Tris pH 8) for 10 minutes at 4°C on a rotating wheel followed by centrifugation to collect nuclei. Nuclei were then resuspended in 1mL LB3 (1mM EDTA pH 8, 0.5mM EGTA pH 8, 10mM Tris pH 8, 100mM NaCl, 0.1% Na-Deoxycholate, 0.5% N-lauroylsarcosine) and sonicated using Bioruptor Pico (Diagenode) in 15mL tubes for 25 cycles of 30 s ON and 30 s OFF pulses in 4°C water bath. All buffers (LB1, LB2 and LB3) were complemented with EDTA free Protease inhibitor cocktail (Roche), 0.2mM PMSF and 1μg/mL Pepstatin just before use. After sonication, Triton X-100 was added to a final concentration of 1% followed by centrifugation at 20000 *rcf.* and 4°C for 10 minutes to remove particulate matter. After taking a 50μl aliquot to serve as input and to analyze fragmentation, chromatin was aliquoted and snap-frozen in liquid nitrogen and stored at −80°C until use in ChIP assays.

Input aliquots were mixed with equal volume of 2X elution buffer (100mM Tris pH 8.0, 20mM EDTA, 2% SDS) and incubated at 65°C for 12 hours for reverse-crosslinking. An equal volume of TE buffer (10mM Tris pH 8 and 1mM EDTA pH 8) was added to dilute the SDS to 0.5% followed by treatment with RNase A (0.2μg/mL) at 37°C for one hour and Proteinase K (0.2μg/mL) for two hours at 55°C. DNA was isolated by phenol:chloroform: isoamylalcohol (25:24:1 pH 8) extraction followed by Qiaquick PCR Purification (QIAGEN, Germany). Purified DNA was then analyzed on a 2% agarose gel or on Bioanalyzer (Agilent, USA) using a High Sensitivity DNA Assay.

Protein-G coated Dynabeads were incubated at 4°C in blocking solution (0.5% BSA in PBS) carrying specific antibodies to prepare beads pre-coated with specific antibody which were then used for ChIP. Sonicated chromatin was added to pre-coated beads and

the mix was incubated overnight at 4°C on a rotating wheel (please refer to the Table S4 for information on specific antibodies and number of cells used for each ChIP). After incubation with chromatin, beads were washed 7 times with Wash buffer (50mM HEPES pH 7.6, 500mM LiCl, 1mM EDTA pH 8, 1% NP-40, 0.7% Na-Deoxycholate, 1X protease inhibitor cocktail) followed by one wash with TE-NaCl buffer (10mM Tris pH 8 and 1mM EDTA pH 8, 50mM NaCl) and a final wash with TE buffer (10mM Tris pH 8 and 1mM EDTA pH 8). Immunoprecipitated chromatin was eluted by two sequential incubations with 50μl Elution buffer (50mM Tris pH 8, 10mM EDTA pH 8, 1% SDS) at 65°C for 15 minutes. The two eluates were pooled and incubated at 65°C for 12 hours to reverse-crosslink the chromatin followed by treatment with RNase A and Proteinase K and purification of DNA as described above for Input samples. Med1 and Ints11 IPs were analyzed by qPCR (Stratagene) following manufacturer recommendations. Purified DNA was quantified with Qubit DS DNA HS Assay (ThermoFisher Scientific, USA).

At lease 1ng of ChIP DNA was used to prepare sequencing library with Illumina ChIP Sample Library Prep Kit (Illumina, USA). After end-repair and adaptor ligation, library fragments were size-selected using E-Gel SizeSelect 2% Agarose Gel (ThermoFisher Scientific, USA) followed by 12 cycles of PCR amplification. Barcoded libraries from different samples were pooled together and sequenced on Illumina HiSeq2000 platform in paired-end sequencing runs.

Total RNA-seq

RNA was extracted from cells using TRIzol Reagent (ThermoFisher Scientific, USA) according to manufacturer's instructions. Any contaminating DNA was digested with rigorous Turbo DNase (ThermoFisher Scientific, USA) treatment according to manufacturer's instruction followed by a second extraction with TRIzol reagent to eliminate traces of contaminants. Purified RNA was quantified with Nanodrop 1000 instrument and quality was assessed using RNA Nano or Pico Assay kit with Bioanalyzer (Agilent Technologies, USA). Only the RNA samples with RIN above 8 were used for sequencing.

For strand-specific sequencing, ribosomal RNA was removed from total RNA with Ribo-Zero rRNA Removal Kit (EpiCenter, USA) according to manufacturer's instructions and depletion of rRNA was confirmed by analyzing the samples on RNA Pico Assay on Bioanalyzer. Libraries were prepared either with ScriptSeq Total RNA Library prep kit (EpiCenter, USA) according to manufacturer's instructions for the comparison of rWT and the 4 mutants shown in Figure 1 or with Small RNA Library Prep Kit (Illumina, USA) using a modified protocol for the data showed in Figure 2 and later as follows: 50ng rRNA depleted total RNA was fragmented to ~150bp by digesting with 1U of RNaseIII (ThermoFisher Scientific, USA) for 10 minutes at 37°C in a 10μl reaction. Fragmentation reaction was stopped by adding 90μl nuclease-free water and quickly adding 350μl RLT buffer from RNeasy Mini Kit (QIAGEN, Germany) followed by purification of fragmented RNA using RNA Cleanup Protocol from this kit however to enhance the recovery of smaller fragments, we added 500μl ethanol instead of recommended 250μl. 20ng RNaseIII fragmented RNA was used as input for ligation of 3′ and 5′ adapters according to Small RNA Library Prep Protocol followed by cDNA synthesis from adaptor ligated RNA and 10 cycles of PCR amplification. However instead of performing a size-selection of agarose gel (as recommended by manufacturer for sequencing of small RNAs e.g., miRNAs), we used 1 volume of Ampure XP Beads (Beckman Coulter, USA) to clean up the amplified library and remove adaptor dimers according to manufacturer's instructions. Purified libraries were then analyzed with HS DNA Assay Kit on Bioanalyzer (Agilent Technologoes, USA) and sequenced on Illumina HiSeq2000 platform.

PolyA RNA-seq

Polyadenylated RNA was isolated from 5μg total RNA sample by two sequential purifications using Dynabeads mRNA Purification Kit (ThermoFisher Scientific, USA) according to manufacturer's instruction. Purified Poly(A) RNA was analyzed on Bioanalyzer using an RNA Pico Assay chip. Sequencing libraries were then prepared using Small RNA Library Prep Kit (Illumina, USA) using the modified protocol as described above for total RNA-seq.

chrRNA-seq

Chromatin associated RNA was isolated from $20x10^6$ cells essentially as described previously by (Nojima et al., 2015) followed by rigorous treatment with TurboDNase. Before library preparation, any contaminating rRNA was removed with Ribo-Zero rRNA Removal Kit and libraries were prepared using Small RNA Library Prep Kit as described above for total RNA-seq.

MNase-Seq

Nucleosomal DNA was obtained by digesting the chromatin with micrococcal nuclease (MNase). For this purpose, $5 x10^6$ Raji cells were resuspended in 50μl Buffer I (150mM sucrose, 80mM KCl, 5mM K2HPO4, 5mM MgCl2, 0.5mM CaCl2, 35mM HEPES pH 7.4) and then permeabilized by adding NP40 to a final concentration of 0.2% while incubating at 37°C for one minute. Then 500μl of Buffer II (150mM sucrose, 50mM Tris pH 8, 50mM NaCl, 2mM CaCl2) was added along with 25 units of MNase enzyme (Roche Diagnostics, USA) and incubated at 37°C for 10 minutes. Reaction was stopped by adding EDTA to a final concentration of 10mM quickly followed by addition of 1.45 mL of SDS Lysis Buffer (1% SDS, 10mM EDTA pH 8, 50mM Tris pH 8). After 10 minutes of incubation at 4°C, 200μl aliquot was processed for extraction of DNA after treatment with RNase A and Proteinase K followed by an extraction with phenol:chloroform:isoamylalcohol (25:24:1). Only those nucleosomal DNA preps were used subsequently where DNA fragments corresponding to mononucleosomal fraction (~147bp) formed at least 70% of all DNA fragments. Sequencing libraries were then prepared with Illumina ChIP Sample Library Prep Kit (Illumina, USA) as described above for ChIP-seq libraries.

II- Bioinformatic Procedures

ChIP-seq Data Processing

For ChIP-seq, raw sequencing reads were aligned to human genome (hg19) using Bowtie2 (Langmead and Salzberg, 2012). Sequence reads that aligned multiple times in genome with equal alignment score, were discarded as well as the duplicate reads

with identical coordinates (sequencing depth taken into account) were discarded to remove potential sequencing and alignment artifacts. Aligned reads were elongated *in silico* using the DNA fragment size inferred from paired-reads or an estimated optimal fragment size for orphan reads using an in-house developed R pipeline named PASHA (Fenouil et al., 2016). These elongated reads were then used to calculate the number of fragments that overlapped at a given nucleotide thus representing an enrichment score for each nucleotide in the genome. Wiggle files representing average enrichment score every 50bp were generated. Sequencing data from Input samples were treated in the same way to generate Input wiggle files. All wiggle files were then rescaled to normalize the enrichment scores to reads per million. Enrichment scores from Input sample wiggle files were then subtracted from ChIP sample wiggle files. This allowed us to remove/reduce the over-representation of certain genomic regions due to biased sonication and DNA sequencing. Besides this, input subtraction also improves the signal/noise ratio especially for ChIPs with low enrichment. Rescaled and Input subtracted wiggle files from biological replicate experiments were then used to generate a wiggle file that represents the average signal from several biological replicates.

RNA-seq Data Processing

Raw sequencing reads were aligned to human genome (hg19) using TopHat2 (Kim et al., 2013). Sequence reads that aligned multiple times in genome with equal alignment score, were discarded. Thanks to strand-specific library prep of RNA samples, we could infer the strand from which the RNA was originally transcribed hence we separated the reads that align to Watson and Crick strands and processed them separately using PASHA (Fenouil et al., 2016) pipeline to generate strand-specific wiggle files. All wiggle files were then rescaled to normalize the enrichment scores to reads per million. Rescaled wiggle files from biological replicate experiments were then used to generate a wiggle file that represents the average strand-specific RNA signal from several biological replicates.

Gene Expression Analysis

Differential Gene Expression (DGE) analysis was performed by using the DESeq package (Anders and Huber, 2010) from Bioconductor. First, HTseq-count program from the HTSeq framework (Anders et al., 2015) was used to count the sequence reads mapping to gene annotations and then these counts were processed using the DESeq package to identify genes that are at least 3-fold (log2) differentially expressed relative to the reference sample.

Identification of genes downregulated due to interference of antisense transcription

We identified all the genes that were downregulated (log2 fold change > 1) in sense transcription as well as the genes that showed upregulation (log2 fold change > 1) in YFFF mutant as compared to rWT with FDR 0.05 and pval < 0.05. Intersection of the two lists of genes gave us the genes that are potentially downregulated due to interference from antisense transcription.

Peak calling

We used wiggle files to detect the genomic regions with enrichment signals beyond background signal. For this purpose, we used *Thresholding* function of the Integrated Genome Browser (IGB) to determine the enrichment score above which we considered a genomic region to be enriched relative to background noise (*Threshold*) as well as minimum number of consecutive enriched bins to be considered an enriched region (*Min.Run*) and finally the minimum gap beyond which two enriched regions were considered to be distinct (*Max.Gap*) (see Table S7 for parameters used). These parameters were then fed to an in-house script that performs peak-calling by using algorithm employed by *Thresholding* function of IGB.

Identification of Active Enhancers

Genomic regions that show simultaneous enrichment with H3K4me1, H3K27ac and Pol II and are at least ± 1500bp away from any annotated gene were considered to be putative active enhancers. To remove any unannotated promoters from our enhancer selection, we filtered out any regions that were more enriched with H3K4me3 as compared to H3K4me1. Method used for this filter has been described previously in Descostes et al., 2014. In identified enhancers, position of the minimum signal of H3K27ac (nucleosome depleted region – NDR) which was closest to location of maximum signal of Pol II was defined as center of the region.

Average Metagene Profiles

To generate average signal profiles, we selected the hg19 genes or identified enhancer regions that do not have any other annotation within 20Kb (Figures 1, 2, and 3), 10kb (Figure 4), 2kb (Figure 5) around boundaries. Removal of the annotations too close to each other is necessary to avoid mixing signals from close-by annotations which can cause misinterpretation of the results. ChIP-seq, MNase-seq and strand-specific RNA-seq values from wiggle files were retrieved with in-house R and Perl scripts for selected genes and enhancer regions. Then we used an algorithm as described previously (Koch et al., 2011) to rescale the genes to same length by interpolating the values on 1000 points and build a matrix on which each column is averaged and resulting values are used to plot average metagene profiles.

RNA RT index

Upstream and downstream RT transcription indices (Figure 2D, 3B, 3H, and S4D) were calculated by dividing average sense (for downstream RT) and antisense (for upstream RT) signal in 20kb region upstream or downstream of the gene with average signal in first half of the corresponding gene body. Asinh transformation was applied to the values for graphical representation.

Pol II read through index

Upstream and Downstream Pol II RT indices (Figures 3E, S4B, and S4C) were calculated by dividing average signal in 10kb region upstream or downstream of the gene respectively with average signal in second half of the corresponding gene body. Asinh transformation was applied to the values for graphical representation.

Pol II pausing score

Pol II pausing score (Figure 4D) was calculated as described earlier (Fenouil et al., 2016). Briefly, the average Pol II ChIP-seq signal in −300bp / +100bp region around TSS was divided by average signal in second half of the corresponding gene body. Asinh transformation was applied to the values for graphical representation.

QUANTIFICATION AND STATISTICAL ANALYSIS

All ChIP-seq, RNA-seq and MNase-seq experiments were performed in at least two biological replicates. Statistical significance of differential metagene profiles was calculated by two sided Wilcoxon test. p values associated to the number of asterisks in figures are described in figure legends. Significance of differential gene expressions were calculated by non-parametric Mann-Whitney test and p values were adjusted for FDR < 0.05. Genes with at least 3-fold change in expression level relative to rWT were considered to differentially regulated.

DATA AND SOFTWARE AVAILABILITY

All high throughput sequencing data used in this study have been deposited at GEO under accession number GSE94330.

The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE [1] partner repository with the dataset identifier PXD008270. The raw western blot figures are accessible online on Medley data through the following link: https://doi.org/10.17632/jkmfb9mnyt.1.