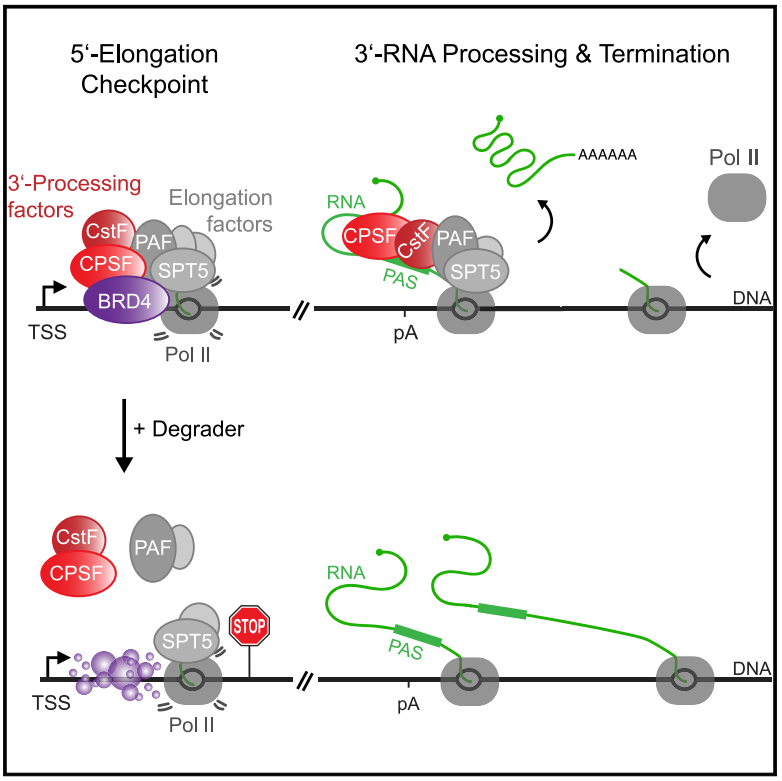
Molecular Cell

Article

A BRD4-mediated elongation control point primes transcribing RNA polymerase II for 3′-processing and termination

Graphical abstract



Highlights

- Acute loss of BRD4 impairs Pol II pause release and prompts readthrough transcription
- BRD4 interacts with 5′-elongation control and 3′-RNA processing factors
- BRD4 helps to recruit 3′-RNA processing factors during a 5′-elongation checkpoint
- BET protein and BRD4 ablations provoke RNA cleavage defects

Authors

Mirjam Arnold, Annkatrin Bressin, Olga Jasnovidova, David Meierhofer, Andreas Mayer

Correspondence

mayer@molgen.mpg.de

In brief

Arnold et al. use rapid BET and BRD4-selective degradation along with quantitative approaches that capture the immediate consequences to reveal direct functions of BRD4 in RNA polymerase II transcription regulation. They provide evidence that BRD4 underlies a general 5′-elongation checkpoint, which primes transcribing RNA polymerase II for 3′ processing and termination.

Arnold et al., 2021, Molecular Cell *81*, 1–15
September 2, 2021
https://doi.org/10.1016/j.molcel.2021.06.026

CellPress

Molecular Cell

CellPress

Article

A BRD4-mediated elongation control point primes transcribing RNA polymerase II for 3′-processing and termination

Mirjam Arnold,[1,2,5] Annkatrin Bressin,[1,3,5] Olga Jasnovidova,[1] David Meierhofer,[4] and Andreas Mayer[1,6,*]
[1]Otto-Warburg-Laboratory, Max Planck Institute for Molecular Genetics, 14195 Berlin, Germany
[2]Department of Biology, Chemistry, and Pharmacy, Freie Universität Berlin, 14195 Berlin, Germany
[3]Department of Mathematics and Computer Science, Freie Universität Berlin, 14195 Berlin, Germany
[4]Mass Spectrometry Facility, Max Planck Institute for Molecular Genetics, 14195 Berlin, Germany
[5]These authors contributed equally
[6]Lead contact
*Correspondence: mayer@molgen.mpg.de
https://doi.org/10.1016/j.molcel.2021.06.026

SUMMARY

Transcription elongation has emerged as a regulatory hub in gene expression of metazoans. A major control point occurs during early elongation before RNA polymerase II (Pol II) is released into productive elongation. Prior research has linked BRD4 with transcription elongation. Here, we use rapid BET protein and BRD4-selective degradation along with quantitative genome-wide approaches to investigate direct functions of BRD4 in Pol II transcription regulation. Notably, as an immediate consequence of acute BRD4 loss, promoter-proximal pause release is impaired, and transcriptionally engaged Pol II past this checkpoint undergoes readthrough transcription. An integrated proteome-wide analysis uncovers elongation and 3′-RNA processing factors as core BRD4 interactors. BRD4 ablation disrupts the recruitment of general 3′-RNA processing factors at the 5′-control region, which correlates with RNA cleavage and termination defects. These studies, performed in human cells, reveal a BRD4-mediated checkpoint and begin to establish a molecular link between 5′-elongation control and 3′-RNA processing.

INTRODUCTION

Transcription elongation has emerged as a highly regulated phase in metazoan gene expression (Noe Gonzalez et al., 2021; Mayer et al., 2017; Zhou et al., 2012). Following transcription initiation, polymerase II (Pol II) pauses in the promoter-proximal region at the majority of mammalian genes (Jonkers et al., 2014; Laitem et al., 2015; Mayer et al., 2015; Rahl et al., 2010). The release of paused Pol II into productive elongation represents a major control point (Core and Adelman, 2019; Chen et al., 2018; Wissink et al., 2019). Much of our knowledge of the mechanisms that underlie pause release is derived from experiments performed *in vitro*. The DRB sensitivity-inducing factor (DSIF), composed of SPT5 and SPT4, and the negative elongation factor (NELF) complex establish the pause with the help of other factors (Vos et al., 2018b; Yamaguchi et al., 1999; Wada et al., 1998). For pause release, the positive transcription elongation factor b (P-TEFb) is required (Peterlin and Price, 2006); it phosphorylates the C-terminal repeat domain (CTD) of Pol II, SPT5, NELF, and others (Sansó et al., 2016; Vos et al., 2018a. Recently, the PAF1 complex (PAF) has also been implicated in pause release Vos et al., 2018a; Yu et al., 2015; Chen et al., 2015). Our picture of factor-specific functions during this early elongation checkpoint that control nascent RNA synthesis in cells is still incomplete.

BRD4 has also been ascribed a contributory role in pause release (Anand et al., 2013; Lovén et al., 2013). BRD4 together with BRD2, BRD3, and BRDT belongs to the bromodomain and extra-terminal domain (BET) protein family (Belkina and Denis, 2012). According to the original view, BRD4 recruits P-TEFb to paused Pol II, leading to pause release (Jang et al., 2005; Yang et al., 2005). Using an optimized BET degrader (dBET6), we previously uncovered that acute loss of BET proteins led to a collapse of transcription elongation, but surprisingly was inconsequential for the recruitment of P-TEFb (Winter et al., 2017). These observations questioned the mechanistic role of BRD4 in pause control. Although the general implication of BRD4 in transcription elongation was recently confirmed (Muhar et al., 2018), the underlying mechanism and whether BRD4 serves additional functions in transcription remained unclear.

Upon successful negotiation of the 5′-elongation checkpoint, Pol II escapes from potential premature termination (Brannan et al., 2012; Krebs et al., 2017; Kamieniarz-Gdula et al., 2019) and enters into processive elongation. At the 3′ end of genes, Pol II transcribes through the polyadenylation (pA) site. The pA signal (PAS) within the nascent RNA is recognized by the cleavage and pA specificity factor (CPSF) (Chan et al., 2014). CPSF contains the endonuclease CPSF73, which cleaves the nascent RNA (Mandel et al., 2006). PAS sensing and cleavage require

Please cite this article in press as: Arnold et al., A BRD4-mediated elongation control point primes transcribing RNA polymerase II for 3′-processing and termination, Molecular Cell (2021), https://doi.org/10.1016/j.molcel.2021.06.026

additional 3′-processing factors including the cleavage stimulation factor (CstF) (Kumar et al., 2019; Sun et al., 2020). Importantly, Pol II continues transcribing past the pA site (Anamika et al., 2013; Mayer et al., 2015; Schwalb et al., 2016). To avoid transcriptional interference with downstream genes, transcription needs to be terminated (Porrua and Libri, 2015; Richard and Manley, 2009; Proudfoot, 2016). This is accomplished by RNA cleavage, which creates an entry site for the exonuclease XRN2, finally releasing Pol II from the chromatin (West et al., 2004; Kim et al., 2004; Fong et al., 2015; Eaton et al., 2018). Clear regulatory links between early transcription elongation and 3′-RNA processing have remained undefined.

Here, we combined BET protein and BRD4-selective degradation in human cells with a quantitative multi-omics approach to uncover direct roles of BRD4 in Pol II transcription. Quantification of nascent transcription under conditions where BRD4 function was impaired uncovers a uniform and global reduction of Pol II transcription. Acute loss of BRD4 impairs Pol II pause release and surprisingly provokes massive readthrough transcription and RNA cleavage defects. We show that BRD4 is required for the recruitment of the 3′-RNA processing machinery during a 5′-elongation control point to enable RNA cleavage and transcription termination. Together, our results identify mechanisms of BRD4-mediated elongation control and uncover a link between a general elongation checkpoint and 3′-RNA processing.

RESULTS

BET degradation uniformly decreases nascent transcription

We previously found that BET proteins play a general role in transcription elongation (Winter et al., 2017). However, the underlying molecular mechanism remained unclear. To uncover the role of BET proteins in Pol II transcription, we developed a protocol for performing native elongating transcript sequencing (NET-seq) that relies on spike-in controls, called SI-NET-seq (Figure 1A). The use of spike-ins allows quantitative comparisons between conditions and can reveal uniform changes in transcript abundance and Pol II occupancy (van de Peppel et al., 2003; Lovén et al., 2013; Orlando et al., 2014). We applied SI-NET-seq to determine the immediate consequences of the acute loss of BET proteins on transcription by treating human MOLT4 cells with the BET degrader dBET6 (Winter et al., 2017). The high correlation between replicate measurements indicated the robustness of this approach (Pearson's correlation r = 0.96; Figure S1A). SI-NET-seq revealed a strong uniform decrease of transcriptionally engaged Pol II upon acute loss of BET proteins (Figure 1B), which was in contrast to the observations obtained with standard NET-seq (Figure 1B). This illustrates the importance of using spike-in methods especially in comparative nascent transcription studies.

Using the SI-NET-seq data, we found that Pol II levels were mainly reduced over the gene-body (Figures 1C and S1B), indicating that BET proteins are required for elongation activation. We observed this global elongation defect at protein-coding and long non-coding RNA (lncRNA) genes, but not at histone genes (Figure 1D). For 80% of active genes, BET degradation led to a significant reduction of transcription elongation that we classified as "BET-sensitive" genes. The other active genes, called "BET-resistant" genes, showed no significant decrease in elongation (Figure 1E). We found that BET-sensitive genes were longer, with a median length of 33 kb compared with 6 kb for BET-resistant genes (Figure 1F). Although we could not detect differences in BRD4 binding at different gene length classes (Figure S1C), we speculated that genes with higher BRD4 levels such as histone genes (Figure S1D) could initially buffer acute loss of BET proteins. Our observation that BRD4 levels at BET-resistant genes were not reduced to the same extent as BET-sensitive genes supported this view (Figure S1E).

SI-NET-seq time-course experiments upon dBET6 exposure showed that the global decrease in elongation is conserved in other cellular contexts (human K562 cells; Figures S1F and S1G) and increased with the treatment time (Figure 1G). This analysis also uncovered a nascent transcriptional wave that moved toward the 3′ end of genes corresponding to transcribing Pol II that already entered into active elongation at the time of degrader treatment (Figures 1H and 1I). These observations indicate that BET degradation perturbs the transition into productive elongation.

We conclude that SI-NET-seq allows direct quantitative comparisons of nascent transcription between conditions, revealing a uniform reduction in elongation upon BET degradation.

BRD4-specific degradation impairs Pol II pause release

Using BET degradation, it remained unknown which of the BET proteins was responsible for the global elongation defect. Based on our observation that BET-sensitive genes were more susceptible to a reduction in BRD4 levels and given the emerging implication of BRD4 in elongation (Jang et al., 2005; Lovén et al., 2013; Winter et al., 2017; Muhar et al., 2018), we speculated that BRD4 loss caused the general elongation defect.

To test this hypothesis, we generated a human cell line expressing an N-terminally degradation-tagged (dTAG) (Nabet et al., 2018) version of BRD4 from its endogenous locus (Figures 2A and S2A). While dTAG-BRD4 levels were reduced compared with untagged BRD4 (Figure S2B), as recently also observed for other tagged proteins (Schick et al., 2021), the dTAG had no effect on the subcellular distribution and cell doubling time, indicating that the tag was not interfering with BRD4 function (Figures S2B and S2C). Treating cells with the degrader (dTAG7) resulted in a fast reduction of both main BRD4 isoforms that are expressed in K562 cells within 30 min and led to a complete loss within 2 h (Figure 2B). Degrader-induced BRD4 depletion was also highly specific, leaving the levels of the other BET proteins, BRD2 and BRD3, unchanged (Figure S2D; Table S1). Within this treatment time, cells were viable with no signs of apoptosis (Figure S2E).

We performed SI-NET-seq upon dTAG7 exposure (Figure S2F). As observed upon pan-BET degradation, BRD4-specific depletion led to a strong reduction of productive elongation at protein-coding and lncRNA genes, but not at histone genes (Figures 2C–2E). Although the overall reduction of productive elongation was not as pronounced as for BET degradation (Figure S2G), it was clearly detectable genome-wide and at single genes (Figures 2C–2E). Notably, the collapse of elongation was accompanied by a significant increase of Pol II levels in the

Please cite this article in press as: Arnold et al., A BRD4-mediated elongation control point primes transcribing RNA polymerase II for 3′-processing and termination, Molecular Cell (2021), https://doi.org/10.1016/j.molcel.2021.06.026

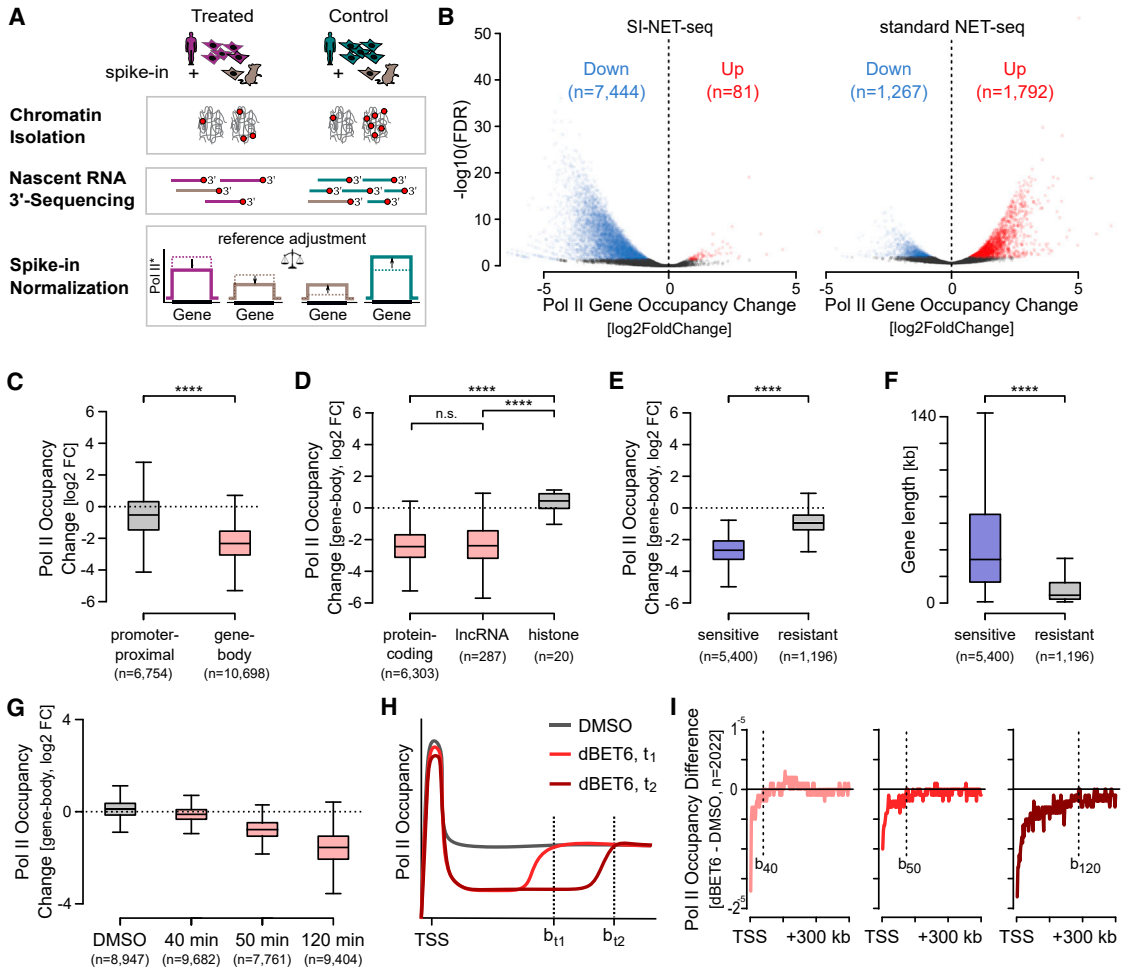


Figure 1. SI-NET-seq informs on the mechanisms of BET-mediated transcription control

(A) Schematic view of the SI-NET-seq approach (STAR Methods). *Pol II occupancy before (dashed lines) and after normalization (solid lines).

(B) Pol II occupancy changes (RLE normalized) upon 2 h dBET6 treatment and for the DMSO control. Significant (FDR < 0.05) occupancy changes are labeled in blue and red.

(C–E) Boxplot quantification of log2 fold-change (FC) from spike-in normalized Pol II occupancy upon 2 h of dBET6 treatment. Significant changes are shown for the promoter-proximal (transcription start site [TSS] to TSS +300 bp) and gene-body (TSS +300 bp to pA site) regions (C). Pol II occupancy changes over the gene-body are indicated for different gene classes (D) and BET-sensitive and -resistant genes (E) (STAR Methods).

(F) Gene length distribution of BET-sensitive and -resistant genes.

(C–F) Wilcoxon rank-sum test; ****$p \leq 5.8e{-}11$; n.s., not significant, p = 0.17.

(G) Boxplot quantification of log2 FC from spike-in normalized Pol II occupancy (SI-NET-seq) over the gene-body at indicated time points for DMSO and upon dBET6 treatment.

(H) Schematic view of the back-edge of the receding Pol II wave determined for each time point ($b_{t1,t2}$) of the SI-NET-seq time-course experiment upon dBET6 exposure.

(I) Difference profiles of the mean Pol II occupancy (SI-NET-seq) between dBET6 treatment and DMSO for the indicated time points. Difference profiles were calculated for long non-overlapping genes (TSS to +300 kb), and 0.1% of the strongest signals was removed.

Data were obtained for human MOLT4 (B–F) and K562 (G–I) cells.

See also Figure S1.

promoter-proximal region (Figures 2C–2E), suggesting that rapid BRD4-specific degradation impairs pause release of Pol II. This observation upon BRD4-specific ablation was a main difference compared with pan-BET degradation where promoter-proximal Pol II levels were slightly decreased (Figure S2G).

Together, these data suggest a general BRD4-mediated elongation checkpoint at the 5′ end of genes.

Rapid BET and BRD4-selective degradation lead to widespread readthrough transcription

Using SI-NET-seq, we next explored whether BET proteins and particularly BRD4 serve additional roles in Pol II transcription. Strikingly, this analysis uncovered that acute loss of BET proteins (dBET6 treatment; Figure 3A) induced massive readthrough transcription of Pol II at the 3′ end of genes (Figures 3B and 3C).

Please cite this article in press as: Arnold et al., A BRD4-mediated elongation control point primes transcribing RNA polymerase II for 3′-processing and termination, Molecular Cell (2021), https://doi.org/10.1016/j.molcel.2021.06.026

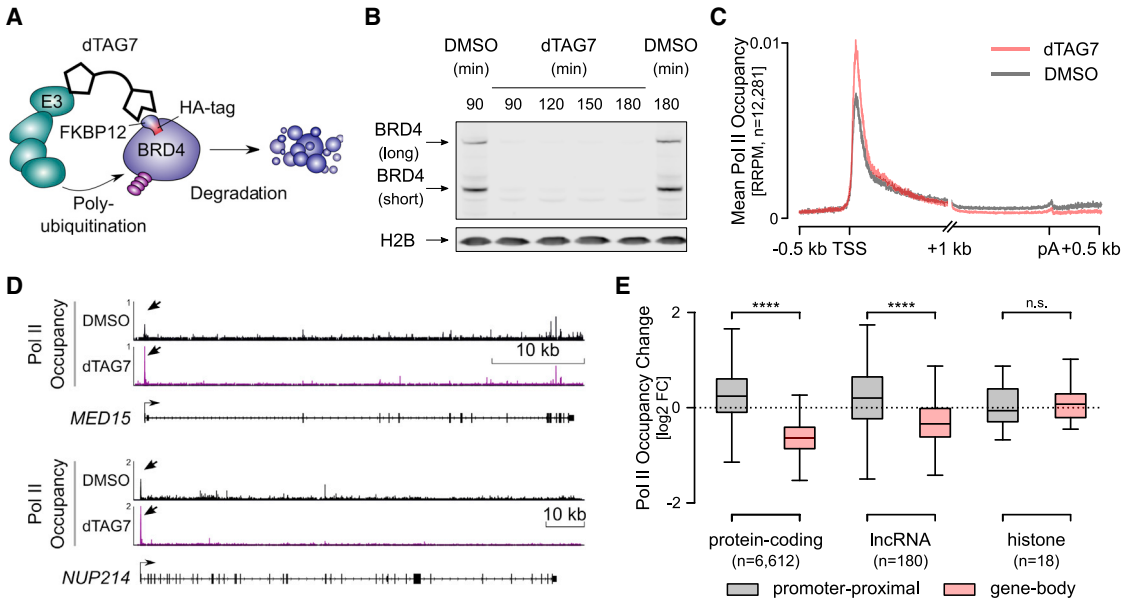


Figure 2. Disruption of Pol II pause release by BRD4-specific degradation

(A) Scheme for BRD4-specific degradation using the PROTAC degrader dTAG7 (Nabet et al., 2018).

(B) Immunoblot for dTAG-BRD4 after dTAG7 (500 nM) treatment and for the non-treated DMSO control shown for the indicated time points. Arrows indicate the long and short BRD4 isoforms expressed in K562 cells. Histone 2B (H2B) served as a loading control.

(C) Meta-gene profile of the mean spike-in normalized Pol II occupancy (reference-adjusted reads per million [RRPM], SI-NET-seq) for actively transcribed genes upon 2 h of dTAG7 treatment and for the DMSO control. Genes of different lengths were scaled in the region 1 kb downstream of the TSS to the pA site, and 0.1% of the strongest signals was removed.

(D) Gene tracks of Pol II occupancy (SI-NET-seq) after 2 h of dTAG7 treatment and for the DMSO control. RRPM values and the genomic position are on the y axis and x axis, respectively. In (C) and (D), biological replicates were pooled for visualization.

(E) Boxplot quantification of spike-in normalized Pol II occupancy changes over the promoter-proximal and gene-body regions for different gene classes upon 2 h of dTAG7 treatment (Wilcoxon rank-sum test; ****$p \leq 2.1e{-}13$; n.s., not significant, $p = 0.62$).

See also Figure S2 and Table S1.

Such a pronounced transcriptional readthrough was not observable upon BET bromodomain inhibition using the inhibitor JQ1 (Figures 3A–3C).

In order to quantify the impact of BET degradation and inhibition on the termination zone, the genomic region in which termination occurs (Fong et al., 2015; Schwalb et al., 2016), we developed the transcriptional "readthrough index" (RTI; Figure 3D). The RTI quantifies changes in the average distance between the pA site and the transcription termination site (Figure 3D). This analysis revealed strong readthrough transcription (RTI $\geq$ 5 kb) of Pol II for 39 and 35% of active genes in human MOLT4 and K562 cells, respectively, upon 2 h of BET degradation. The median length of the 3′ shift upon BET ablation was 3.2 kb in MOLT4 (Figure 3E; Table S2) and 3.0 kb in K562 cells (Figure S3A). Widespread readthrough transcription was clearly visible already after 40 min of degrader treatment when 40% of BET proteins were degraded and substantially increased with longer treatment times (Figure 3F). Exposure to JQ1 led to a mild but significant readthrough (median RTI = 512 bp; Figure 3E). A likely explanation for the difference in the extent of readthrough was that JQ1 treatment led only to an incomplete dissociation of BRD4 from the chromatin (Figure S3B).

Interestingly, BRD4-specific ablation (dTAG7 treatment; Figure 3A) phenocopied the transcriptional readthrough upon pan-BET degradation (Figures 3G and 3H). The extent of the readthrough was similar (Figure 3I). Although fewer genes classified as readthrough genes, they strongly overlapped with those upon BET degradation (Figure S3C). The readthrough varied between different genes (Figure 3J) and increased with the distance between genes (Figure 3K).

Given the accumulating evidence that transcription termination is controlled differently at distinct gene classes (Porrua and Libri, 2015; Richard and Manley, 2009), we next analyzed whether BRD4-specific degradation results in class-specific termination defects. We observed the strongest termination defects for protein-coding and lncRNA genes (protein-coding genes: median RTI = 1.5 kb; lncRNA genes: median RTI = 1.2 kb; Figure S3D). We found no evidence for readthrough transcription at gene classes with non-canonical 3′-RNA processing and termination including histone and small nucleolar RNA (snoRNA) genes (Figure S3D). Specific Gene Ontology (GO) terms were not enriched for readthrough genes (dTAG7 and dBET6), suggesting that BET proteins and BRD4 in particular play a general role in transcription termination. Furthermore, no readthrough was observed for antisense transcription upon acute BRD4 ablation (Figure S3E).

Together, these data indicate that BRD4 is required for transcription termination at protein-coding and lncRNA genes.

Please cite this article in press as: Arnold et al., A BRD4-mediated elongation control point primes transcribing RNA polymerase II for 3′-processing and termination, Molecular Cell (2021), https://doi.org/10.1016/j.molcel.2021.06.026

CellPress

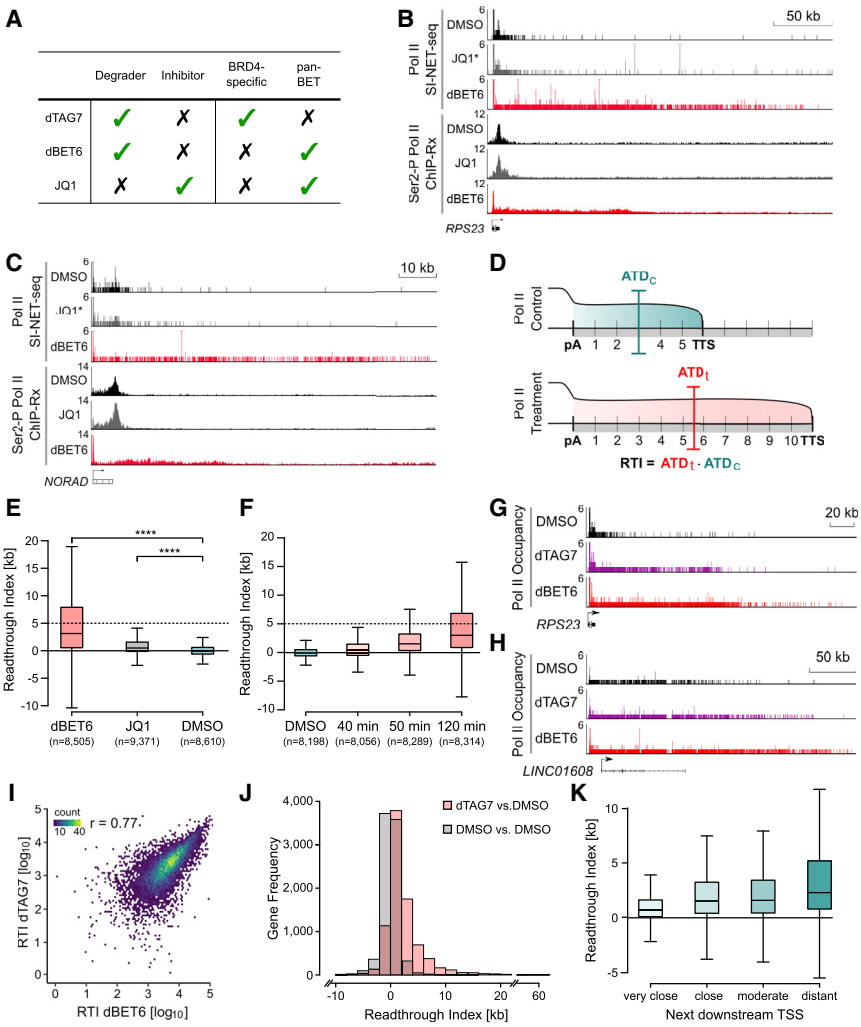


Figure 3. BET and BRD4-specific degradation induce global readthrough transcription

(A) Overview of the compounds utilized in this study to probe pan-BET and BRD4-specific functions.

(B and C) Gene tracks of Pol II and Ser2-P Pol II (ChIP-Rx) occupancies for DMSO and after 2 h treatment with JQ1 and dBET6 at a protein-coding (B) and at a lncRNA (C) gene in MOLT4 cells. The y axis shows SI-NET-seq (*standard NET-seq for JQ1) and ChIP-Rx signals (reads/bp). Biological replicates were pooled for visualization. The NET-seq data (JQ1) and ChIP-Rx data were re-analyzed from Winter et al. (2017).

(D) Schematic view of the "readthrough index" (RTI) calculation. The RTI is defined as the difference of the average termination distance (ATD; STAR Methods) of Pol II between the control measurement and upon treatment. pA, polyadenylation site; TTS, transcription termination site.

(E) Boxplot quantifications of RTI calculations for drug treatments and for the DMSO control in MOLT4. Wilcoxon rank-sum test: ****p < 2.2e−16.

(F) Boxplot quantification of RTI calculations upon dBET6 treatment for indicated time points (SI-NET-seq).

(G and H) Gene tracks of Pol II occupancies (SI-NET-seq) for DMSO and upon 2 h exposure with dTAG7 or dBET6 at a protein-coding (G) and at a lncRNA (H) gene. The y axis shows SI-NET-seq signals (reads/bp).

(I) Scatterplot comparison of the RTI between 2 h of dBET6 and dTAG7 treatments (Pearson's correlation, r = 0.77).

(J) Histogram of the RTI distribution upon 2 h of dTAG7 treatment and between two control measurements (DMSO).

(K) Boxplot quantification of RTI calculations for gene classes with different distances between genes. The distance refers to the region from the pA site of the upstream gene to the TSS of the next active downstream gene (very close, <39 kb; close, 39–135 kb; moderate, 135–465 kb; and distant, >465 kb).

(F–K) Data were obtained for K562 cells.

See also Figure S3 and Table S2.

Readthrough transcription correlates with expression of downstream genes

We next investigated the functional consequences of the massive readthrough transcription. We speculated that the widespread transcriptional readthrough impacts the expression of downstream genes. To explore this, we re-analyzed our previous RNA sequencing (RNA-seq) data (Winter et al., 2017), confirming that the majority of genes (70%) was downregulated after dBET6 treatment but that a subset of genes was significantly upregulated (1,591 genes; Figure S4A). Notably, we found that 91% of upregulated genes were accompanied by upstream readthrough genes (Figure 4A). Readthrough transcription thus provided a potential explanation for this unexpected upregulation. Close inspection of these examples revealed that the transcripts were spliced (Figures 4B and 4C). For convergently oriented genes, we found that readthrough transcription correlated with reduced expression of a set of downstream genes (Figure 4D).

We then asked whether readthrough genes were characterized by specific features. Interestingly, we found that genes with higher RTI scores had significantly higher transcript levels (Figure 4E), increased levels of transcriptionally engaged Pol II (Figure 4F), and a longer median length (Figure 4G). Additionally, we found that readthrough genes had a higher AT content downstream of the pA site, whereas genes without transcriptional readthrough had a higher GC content in this region (Figure 4H). Readthrough genes also had significantly lower H3K79me2 levels over the gene-body (Figure S4B). The number of alternative pA sites and H3K36me3 levels were not significantly linked to readthrough genes (Figures S4C and S4D).

Together, these data show that the transcriptional readthrough correlates with changes in the expression state of neighboring genes and that readthrough genes exhibit distinct features.

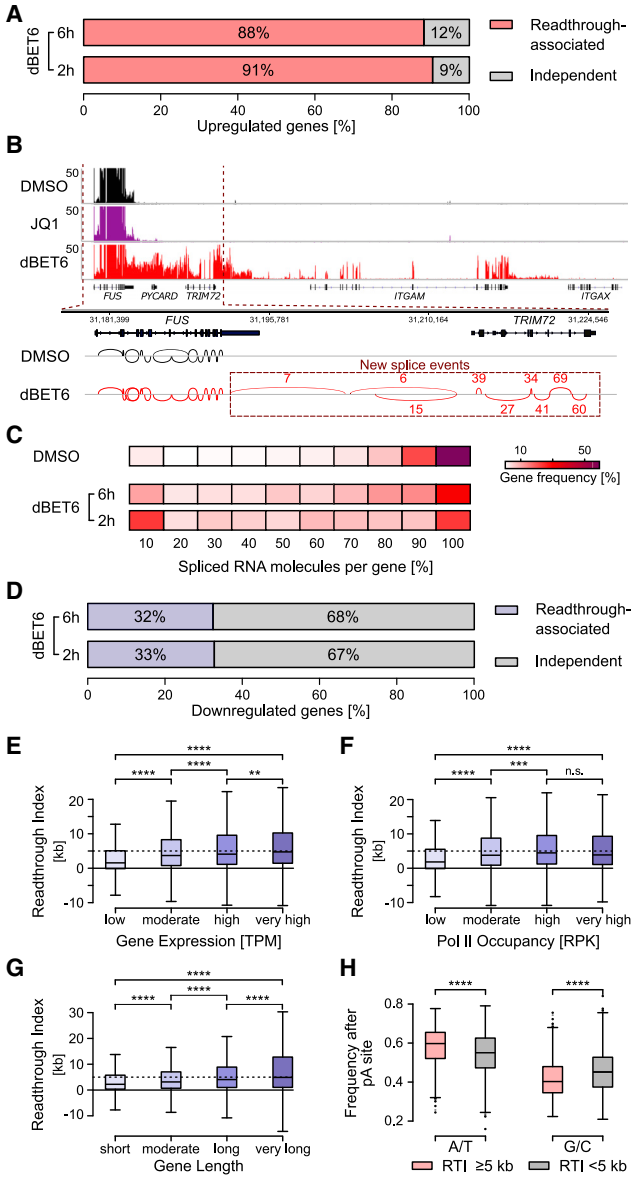


Figure 4. Functional consequences of readthrough transcription for downstream genes

(A) Fraction of differentially upregulated genes associated with readthrough transcription of the upstream gene after 2 and 6 h of dBET6 treatment.

(B) Gene track of total RNA levels (RNA-seq) after 6 h of JQ1 and dBET6 treatment and for the DMSO control. Genes located downstream of *FUS* were likely activated by readthrough transcription with new splicing events (>5 reads) as shown by the Sashimi plot. The y axis depicts RNA levels (reads/bp). Biological replicates were pooled for visualization.

(C) Heatmap of the distribution of spliced RNA molecules per gene across all active genes in DMSO and for readthrough activated genes (STAR Methods) after 2 and 6 h of dBET6 treatment.

(D) Fraction of differentially downregulated genes associated with readthrough transcription of the upstream gene over a convergently oriented downstream gene after 2 and 6 h of dBET6 treatment.

(E–G) Boxplot quantifications for RTI dependencies on 4-quantiles of total gene expression (transcript per kilobase million [TPM] from RNA-seq) (E), Pol II occupancy (reads per kilobase [RPK] from SI-NET-seq) (F), and gene length (kb) (G) upon 2 h of dBET6 treatment.

(H) Boxplot quantification of DNA sequence composition in the genomic region downstream of the pA site of readthrough versus non-readthrough genes. A/T, adenine and thymine; G/C, guanine and cytosine. Wilcoxon rank-sum test one tailed (E–G) and two tailed (H): ****$p \leq 7.2e{-}05$; ***$p = 2.5e{-}04$; **$p = 5.4e{-}03$; n.s., not significant, $p = 0.91$.

(A–H) Data were obtained for MOLT4 cells.

See also Figure S4.

Please cite this article in press as: Arnold et al., A BRD4-mediated elongation control point primes transcribing RNA polymerase II for 3′-processing and termination, Molecular Cell (2021), https://doi.org/10.1016/j.molcel.2021.06.026

CellPress

BET and BRD4 degradation provoke 3′-RNA cleavage defects

Since we observed readthrough transcription upon BET and BRD4-selective degradation mainly at PAS-containing genes, we asked whether 3′-RNA cleavage was defected. Application of an RNA cleavage assay revealed a cleavage defect for two genes (*MBNL1* and *PVT1*) with strong transcriptional readthrough, but not for a control gene (*CDC42*) for which no readthrough was detected upon acute loss of BET proteins (Figures 5A and S5A). This cleavage defect was also detected upon BRD4-specific degradation (Figure S5B).

We next asked whether impaired RNA cleavage occurred globally at readthrough genes. To address this question, we used nanopore sequencing to increase the fraction of reads that span cleavage sites by obtaining long reads. Current protocols for nanopore sequencing only allow the analysis of polyadenylated RNAs. Since the readthrough transcripts that we detected were most likely lacking poly(A) tails due to cleavage defects, we developed an approach to sequence RNAs through the nanopore irrespective of the pA status, similarly to a recently published method (Drexler et al., 2020). Our method, called nascONT-seq, used a short pulse with the uridine analog 4-thiouridine (4sU) to label newly synthesized RNA, chromatin fractionation to enrich for nascent readthrough transcripts, and poly(A) tailing to also capture RNAs that lack a poly(A) tail (Figure 5B).

The correlation between replicate measurements was high, indicating the robustness of this approach (Figure S5C). Application of nascONT-seq confirmed readthrough transcription at selected genes (Figure 5C) and genome-wide upon BET degradation (Figure 5D). Notably, we obtained a large fraction of reads that span cleavage sites, extending our observation that BET degradation perturbed RNA cleavage (Figure 5E). Interestingly, the RTI was significantly increased for genes with a reduced RNA cleavage efficiency (Figure 5F), defined as the ratio of reads that are cleaved versus all reads spanning the pA site. nascONT-seq revealed that the median transcript length increased from 689 to 938 bp upon acute loss of BET proteins (Figure 5G). We could detect extended transcripts originating from 994 genes representing 76% of detected genes (Figure S5D), with the most extended transcript (7.5 kb) for *CREB3L2* (Figure S5E). It is likely that the observed increase in transcript length and the fraction of genes affected were underestimates.

Together, these findings suggest that BET proteins are required for 3′-RNA processing of PAS-containing genes in human cells.

Acute loss of BRD4 disrupts the recruitment of the 3′-processing machinery

Previous work showed that a disruption of 3′-processing and termination factors can cause cleavage defects and readthrough transcription (Porrua and Libri, 2015; Richard and Manley, 2009). We therefore tested whether BRD4 degradation perturbed chromatin localization of 3′-processing factors. Chromatin mass spectrometry (chromatin-MS) revealed that of 964 reproducibly detected proteins, 76 were immediately displaced from the chromatin upon acute loss of BRD4 (p value < 0.05; Figures 6A and S6A; Table S3). Among the strongest hits were factors that have been implicated in 3′-RNA processing (CPSF and CstF) and in transcription elongation including SPT5, SPT6, and PAF (Figures 6A and 6B; Table S3). Consistently, our GO term analysis revealed related significant GO terms (Figure 6C). Similar results were obtained upon pan-BET degradation (Figure S6B). Other factors that have been implicated in elongation control including CDK12, Integrator, Mediator, NELF, P-TEFb, and TFIID were not depleted from the chromatin upon acute BET or BRD4 ablation (Figure S6C; Table S3).

To investigate whether the recruitment of the 3′-processing machinery was perturbed at readthrough genes, we performed chromatin immunoprecipitation with reference exogenous genome (ChIP-Rx) experiments for different subunits of CPSF (CPSF73 and FIP1) and CstF (CstF64). ChIP-Rx relies on spike-in controls that allow quantitative comparisons between samples (Orlando et al., 2014; Winter et al., 2017). This analysis revealed peak occupancy levels of CPSF and CstF subunits at the 5′ and 3′ ends of active genes (Figures 6D and S6D). Moreover, the occupancy of CPSF and CstF significantly decreased over the gene-body upon BRD4-selective ablation, indicating their presence during elongation (Figure S6E). These findings suggest that in human cells CPSF and CstF are already recruited during an early phase of transcription and are in line with previous observations (Dantonel et al., 1997; Glover-Cutter et al., 2008; Davidson et al., 2014; Kamieniarz-Gdula et al., 2019).

Upon BRD4 degradation, we observed a strong decrease of CPSF and CstF occupancy at the 5′ end and downstream of the pA site of genes (Figures 6E and S6F–S6H). This decrease at the 5′ region was significantly stronger than the reduction of Pol II upon acute loss of BRD4, suggesting that the recruitment of 3′-processing factors was impaired (Figure 6F). The recruitment defect was more pronounced at genes with perturbed 3′-RNA cleavage (Figures 6F, S6I, and S6J). These findings suggest impaired recruitment of 3′-processing factors as a likely cause for the cleavage defect and the transcriptional readthrough.

Given the link between Pol II CTD serine 2 phosphorylation (Ser2-P) and 3′-RNA processing (Buratowski, 2009; Bentley, 2014), and the recent observation that BET degradation led to a reduction of Ser2-P over the gene-body (Winter et al., 2017; Muhar et al., 2018), we asked whether this decrease may have caused the recruitment defects. ChIP-Rx of Pol II Ser2-P upon acute BRD4 ablation revealed (1) an increase in Ser2-P levels at promoter-proximal regions similar to BET degradation (Figure S6K), (2) that the general decrease of Ser2-P levels over the gene-body and the 3′ end mainly followed the overall reduction of Pol II (Figure S6L), and (3) that readthrough Pol II was CTD Ser2 phosphorylated (Figure S6L). Although these findings questioned a direct role of Ser2-P in the initial recruitment of 3′-processing factors at the 5′ end of genes, they do not rule out that Ser2-P is required for the maintenance of 3′-processing factors during elongation.

BRD4 interacts with 5′-elongation control and 3′-processing factors

To gain further insights into how BRD4 mechanistically couples 5′-elongation control with 3′-RNA processing, we first performed BRD4-specific immunoprecipitation coupled with MS (IP-MS; Figure S7A) (Mohammed et al., 2016). The IP-MS analysis revealed 379 significant interactors (false discovery rate [FDR] $<$ 0.05) of BRD4 (Figure 7A; Table S4). Apart from known BRD4

Please cite this article in press as: Arnold et al., A BRD4-mediated elongation control point primes transcribing RNA polymerase II for 3′-processing and termination, Molecular Cell (2021), https://doi.org/10.1016/j.molcel.2021.06.026

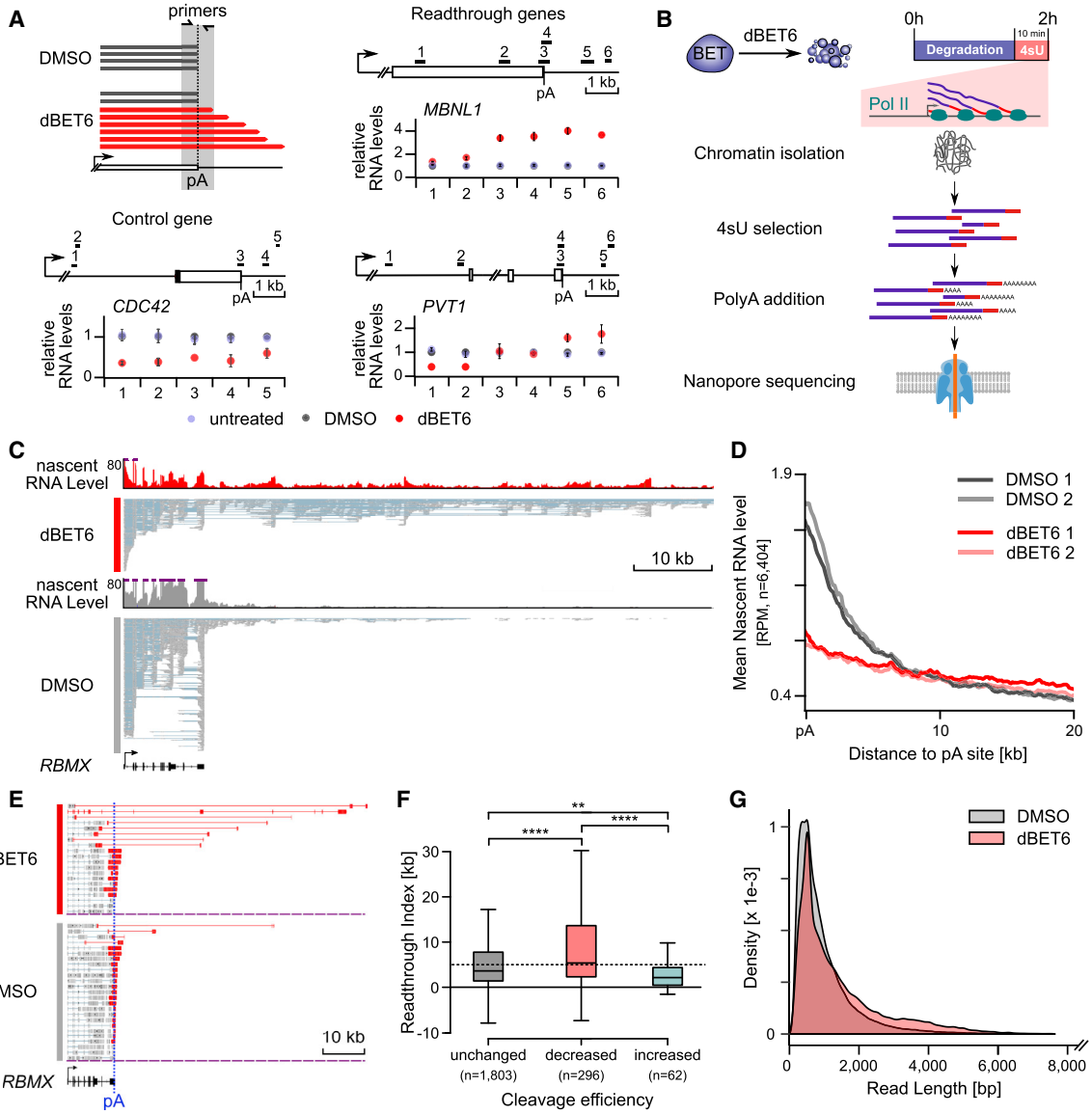


Figure 5. BET degradation perturbs 3′-RNA cleavage at selected genes and genome-wide

(A) Scheme and quantification of 3′-RNA cleavage efficiency at selected readthrough genes and for a control gene showing no transcriptional readthrough upon 2 h dBET6 treatment using qRT-PCR. The positions of PCR amplicons are indicated above the gene. Error bars show the standard deviation between three biological replicate measurements.

(B) Schematic view of the main steps of the nascONT-seq approach.

(C) Gene track of nascent RNA levels and of individual transcripts as obtained by nascONT-seq after 2 h of dBET6 treatment and for DMSO. The y axis depicts the nascent RNA level (reads/bp). Biological replicates were pooled for visualization.

(D) Quantification of nascent RNA levels over the transcription termination zone of active genes after 2 h dBET6 treatment and for DMSO using nascONT-seq.

(E) Individual transcripts (reads) spanning the *RBMX* gene after 2 h dBET6 treatment and for DMSO. Transcripts spanning the pA site are shown in red. Biological replicates were pooled for visualization.

(F) Boxplot quantification for RTI values of genes with different cleavage efficiency changes upon 2 h dBET6 treatment (Wilcoxon rank-sum test: ****$p < 2.3e{-}06$; **$p = 2.8e{-}03$).

(G) Length distribution of nascent transcripts for DMSO (n = 223,355) and upon 2 h dBET6 treatment (n = 428,195) as determined by nascONT-seq. Biological replicates were pooled for visualization.

(A–G) Data were obtained for K562 cells.

See also Figure S5.

Please cite this article in press as: Arnold et al., A BRD4-mediated elongation control point primes transcribing RNA polymerase II for 3′-processing and termination, Molecular Cell (2021), https://doi.org/10.1016/j.molcel.2021.06.026

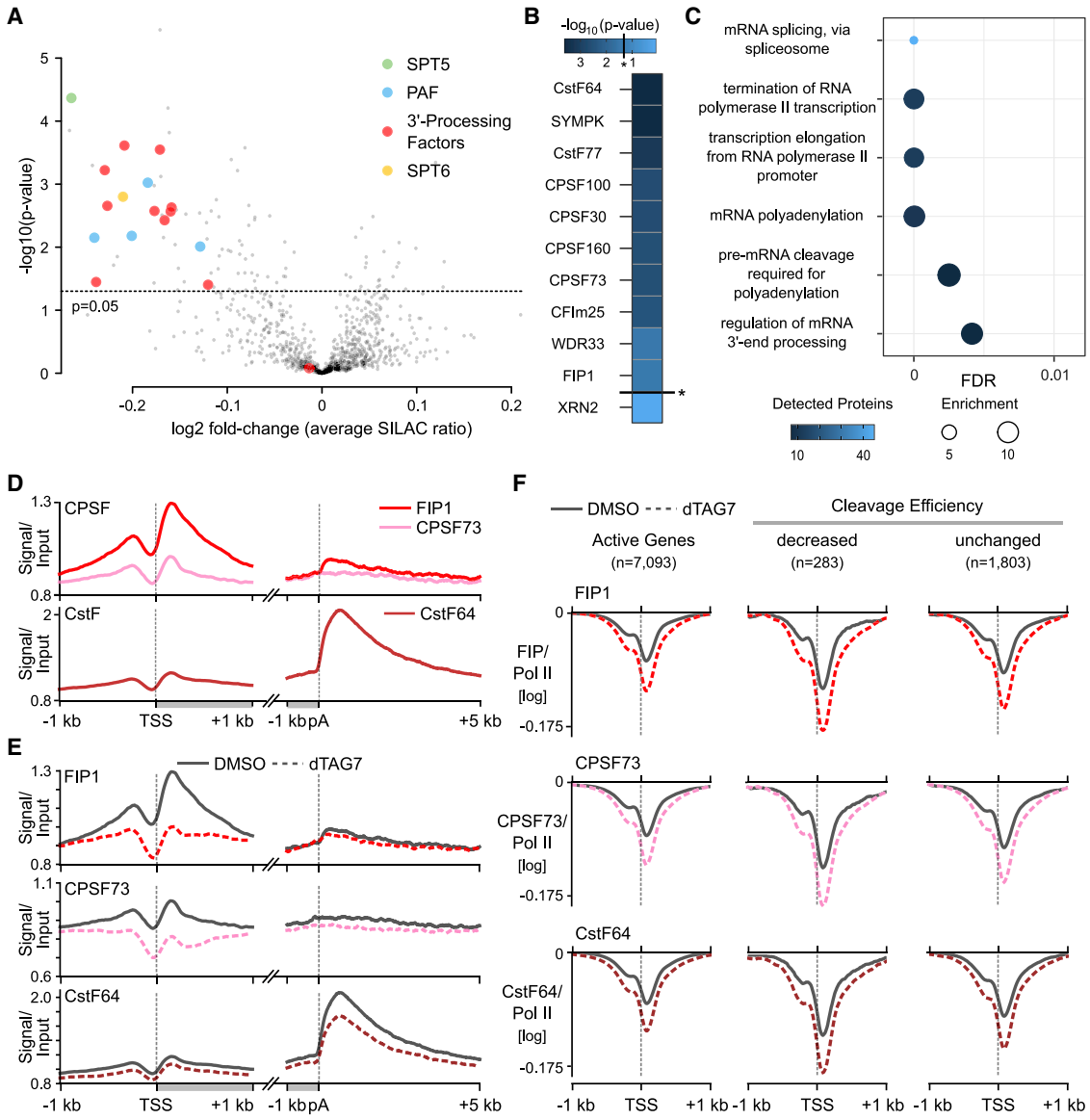


Figure 6. BRD4 degradation impairs the recruitment of the 3′-processing machinery

(A) Changes of chromatin composition upon 2 h dTAG7 treatment as determined by quantitative chromatin-MS.

(B) Heatmap of 3′-RNA processing and termination factors detected by chromatin-MS upon dTAG7 treatment ranked by p value. *Threshold of −log10(0.05).

(C) Summary of significant (FDR < 0.01) Gene Ontology (GO) terms for proteins that were significantly depleted from the chromatin after 2 h of dTAG7 exposure.

(D–F) Meta-gene profiles of occupancy levels (ChIP-Rx) for 3′-RNA processing factors at actively transcribed genes after 2 h dTAG7 treatment and for DMSO. The y axis depicts FC over matched input (D and E) and Pol II-normalized enrichment (F) values.

See also Figure S6 and Table S3.

binders such as ATAD5, BICRA, CHD4, NSD3 (Rahman et al., 2011; Zhang et al., 2016), and P-TEFb (CDK9 and Cyclin T1) (Schröder et al., 2012), other key players of the 5′-elongation transition including PAF, DSIF, and NELF were among the top-ranking interacting proteins (Figures 7A and S7B). Notably, we also detected a set of 3′-RNA processing factors including several subunits of CPSF and CstF (Figures 7A and S7B) as BRD4 interactors. Consistently, the GO term analysis revealed a significant overrepresentation of "mRNA 3′ end processing" (FDR = 9e−03), "transcription elongation from RNA polymerase II promoter" (FDR = 4.3e−05), and related terms (Table S5).

We next integrated our BRD4 IP- and chromatin-MS data to reveal a core BRD4 interaction network (Figure 7B). This core network consisted of significant BRD4 interactors (FDR ≤ 0.005) that were immediately displaced from the chromatin upon BRD4 degradation. Interestingly, it comprised subunits of CPSF (CPSF160), CstF (CstF77), DSIF (SPT5), and PAF (CDC73; Figure 7B). The network also included splicing and DNA repair factors

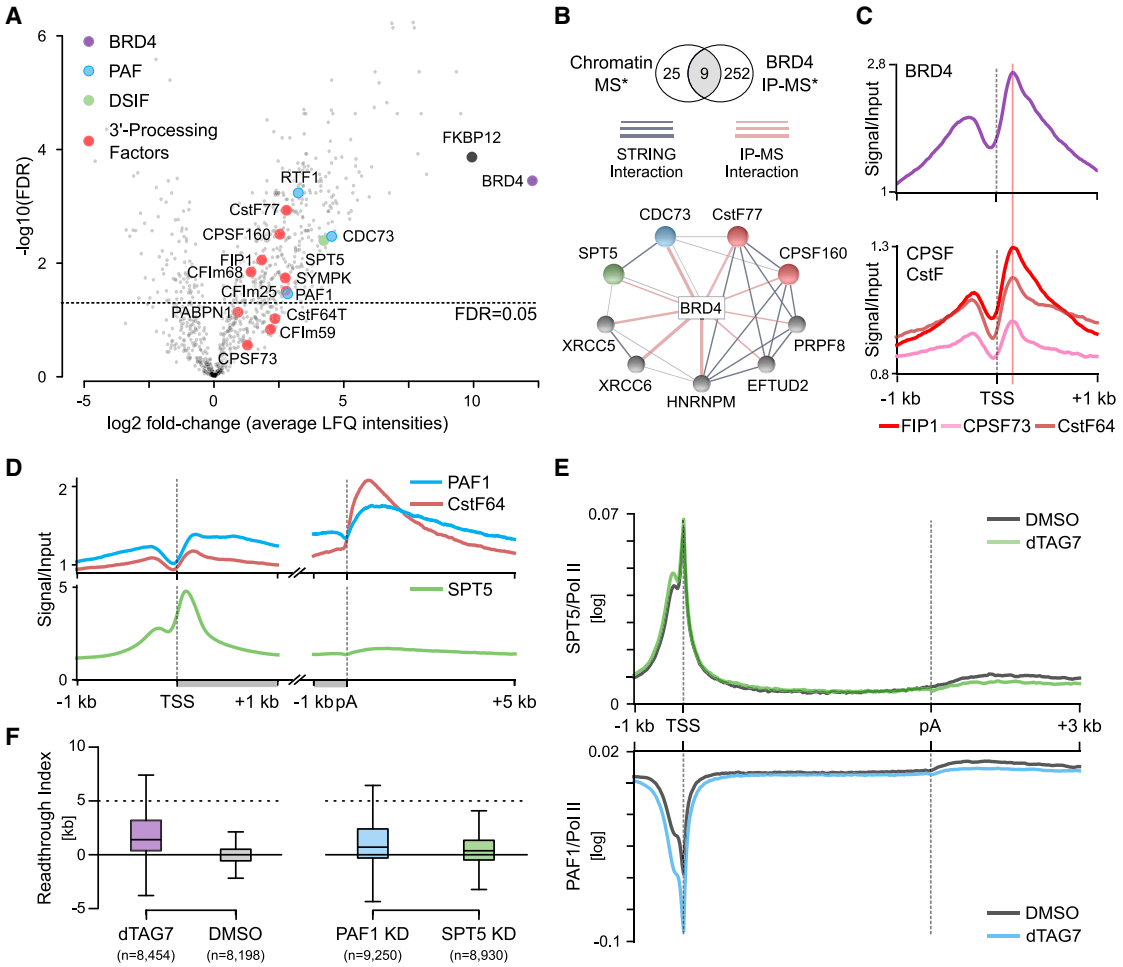


Figure 7. A BRD4 core interaction network links 5′-elongation control and 3′-RNA processing

(A) Interactome of BRD4 as identified by IP-MS. The significance threshold (FDR < 0.05) is indicated by the dashed line.

(B) Core-interaction network of highly significant BRD4 interaction partners (IP-MS, *FDR < 0.005), which are significantly (chromatin-MS, *p < 0.005) displaced from the chromatin after 2 h dTAG7 treatment. New and known interactions are marked by red (BRD4 IP-MS) and gray (STRING database) (Szklarczyk et al., 2019) edges between proteins, respectively.

(C) Meta-gene profiles of input-normalized occupancy levels for BRD4 (ChIP-seq) (Davis et al., 2018) and CPSF and CstF subunits (ChIP-Rx) at actively transcribed genes. The location of occupancy peaks is marked by a red line.

(D) Meta-gene profiles of input-normalized occupancy levels (ChIP-Rx) for PAF1, CstF64, and SPT5 at actively transcribed genes.

(E) Meta-gene profiles of Pol II-normalized SPT5 and PAF1 occupancy levels (ChIP-Rx) at actively transcribed genes after 2 h of dTAG7 treatment and for DMSO.

(A–E) Data were obtained for K562 cells.

(F) Boxplot quantification for the RTI values (STAR Methods) upon BRD4 degradation and PAF1 and SPT5 depletions. RTIs for BRD4 degradation were calculated from SI-NET-seq data obtained for K562. RTIs for PAF1 and SPT5 depletions were calculated from global run-on (GRO)-seq data available for HCT116 and mouse B cells, respectively (Chen et al., 2015; Fitz et al., 2020).

See also Figure S7 and Tables S4 and S5.

(Figure 7B), which is in line with recent studies linking BET proteins to splicing and DNA repair (Uppal et al., 2019; Li et al., 2018).

Since the elongation factor SPT6 showed the strongest consistent reduction upon dBET6 and dTAG7 treatment (Figures 6A and S6B), but was not reliably detected in IP-MS replicates, we performed native IP experiments for BRD4 and SPT6 to clarify potential interactions. Indeed, both factors co-precipitated reciprocally, suggesting an interaction (Figure S7C).

In agreement with the finding that BRD4 interacts with 3′-RNA processing and 5′-elongation control factors, we found that BRD4 co-localizes with these factors in the 5′-control region genome-wide (Figures 7C, S1C, and S7D).

Together, these data suggest that BRD4 can link 5′-elongation and 3′-RNA processing through its functional interactions with 5′- and 3′-control factors.

BRD4 interactors contribute to 3′-processing defects

Despite the evidence that BRD4 recruits 3′-RNA processing factors during the general 5′-elongation control point, we explored whether the observed 3′ end defects were partially mediated

Please cite this article in press as: Arnold et al., A BRD4-mediated elongation control point primes transcribing RNA polymerase II for 3′-processing and termination, Molecular Cell (2021), https://doi.org/10.1016/j.molcel.2021.06.026

through BRD4 interactors. We focused on PAF (CDC73) and SPT5 as they appeared as core BRD4 interactors and based on their emerging roles in 3′-RNA processing in yeast (Nordick et al., 2008; Nagaike et al., 2011; Ellison et al., 2019; Lindstrom et al., 2003; Glover-Cutter et al., 2008; Mayer et al., 2012; Baejen et al., 2017), and of PAF in alternative pA site usage in mammals (Yang et al., 2016).

We performed ChIP-Rx to investigate whether acute loss of BRD4 interfered with the recruitment of PAF (PAF1) and DSIF (SPT5) to transcribing Pol II. The occupancy profile of PAF1, but not of SPT5, had striking similarities with profiles of 3′-processing factors, with maximal occupancy levels at the promoter-proximal region and downstream of the pA site (Figure 7D). A re-analysis of data for other PAF subunits confirmed this finding (Figure S7E). This genome-wide co-localization suggested a potential interaction. Consistently, native IP experiments for several PAF subunits revealed interactions with 3′-processing factors (Figure S7F).

Interestingly, when we normalized PAF1 and SPT5 occupancy signals to the corresponding Pol II levels, the occupancy of PAF1 was, similarly to 3′-processing factors, more strongly reduced at the 5′-control region (Figures 7E, S7G, and S7H), indicating that PAF recruitment was perturbed. For PAF1, the occupancy was also significantly more reduced over the gene-body and at the termination zone as compared to Pol II levels (Figure S7I). For SPT5, a reduced occupancy could be observed at the termination zone (Figures 7E, S7G, S7H, and S7J).

We next asked whether the loss of PAF1 or SPT5 could induce readthrough transcription. To address this question, we re-analyzed genome-wide Pol II occupancy data upon PAF1 or SPT5 knockdown (Chen et al., 2015; Fitz et al., 2020). Indeed, these analyses revealed readthrough transcription upon both knockdowns (Figure 7F).

Together, our findings suggest that the observed 3′ defects are partially mediated through BRD4 interactors.

DISCUSSION

By combining rapid BET and BRD4-specific degradation with a quantitative multi-omics approach, able to capture the immediate effects, our integrative analysis provides a comprehensive picture of BRD4 in the regulation of Pol II transcription. We show that BRD4 underlies a general 5′-regulatory hub that controls productive transcription elongation and has far-reaching consequences for 3′-RNA processing and transcription termination.

BRD4-selective ablation compared with pan-BET degradation revealed a main difference in the functional impact on the 5′-elongation checkpoint. In contrast to BET degradation, acute loss of BRD4 led to a significant accumulation of Pol II in the promoter-proximal region indicative for impaired pause release. Our observation of a nascent transcriptional wave corresponding to transcribing Pol II that had passed the 5′ checkpoint at the time of degrader treatment supported this view. The impaired pause release explained the global collapse in elongation, as also observed previously (Winter et al., 2017; Muhar et al., 2018). The differences upon BRD4-specific and pan-BET protein degradation imply potential roles for BRD2 and BRD3 in 5′-elongation control that are distinct from BRD4. This is in line with the emerging view that BET proteins can have different roles in gene regulation (Cheung et al., 2017; Hsu et al., 2017; Zheng et al., 2021). To clarify these roles represents an interesting subject for future investigations.

Our integrative approach informs on the underlying mechanism of elongation control by BRD4. We identified PAF and DSIF (SPT5) as core BRD4 interactors. This finding is consistent with prior work showing that BRD4 co-precipitates with PAF (Yu et al., 2015) and DSIF (Winter et al., 2017). Notably, we found that BRD4 is required for the recruitment of PAF during the 5′-elongation checkpoint. PAF is an integral component of the activated transcription elongation complex facilitating transcription elongation *in vitro* and *in vivo* (Vos et al., 2020; Yu et al., 2015; Hou et al., 2019; Chen et al., 2009; Kim et al., 2010), although an antagonistic role was also described (Chen et al., 2015). In addition, we identified an interaction between BRD4 and SPT6, another subunit of the activated elongation complex Vos et al., 2018a; Endoh et al., 2004) essential for elongation (Li et al., 2021; Nojima et al., 2018). These findings support the view that BRD4 is required for the assembly of a functional Pol II elongation complex capable of elongation.

Unexpectedly, BET and BRD4-selective degradation impaired 3′-RNA cleavage and transcription termination. As a potential underlying mechanism, we identified the perturbed recruitment of integral components of the general 3′-RNA processing machinery (CPSF and CstF) upon acute loss of BRD4. These alterations likely disrupted PAS sensing and RNA cleavage and as a consequence may have perturbed the entry of the torpedo termination factor XRN2 (Fong et al., 2015; Eaton et al., 2018), inducing widespread readthrough transcription. Our model is also consistent with the observation that CPSF depletion provokes widespread transcriptional readthrough (Nojima et al., 2015; Eaton et al., 2020; Lykke-Andersen et al., 2021) and with the previous finding that RNA cleavage is a requirement for efficient transcription termination at PAS-containing genes (Eaton and West, 2020).

Mechanistically, our data suggest that BRD4 links 3′-RNA processing to the 5′-elongation checkpoint through the assembly and release of a 3′-processing competent elongation complex. Several lines of evidence suggest a direct role of BRD4. First, the recruitment of 3′-processing factors to the 5′ end of target genes is impaired upon BRD4 degradation. Second, acute loss of BRD4 provokes an immediate dissociation of 3′-processing and selected elongation factors from the chromatin. Third, BRD4 interacts with 3′-processing and 5′-elongation control factors. Finally, BRD4 co-localizes with 3′-RNA processing and pausing factors at promoter-proximal regions genome-wide.

Despite a direct link between BRD4 and 3′-RNA processing we also found that the impact of BRD4 on 3′-processing and termination can partially be mediated through its interactors. Multiple lines of evidence suggest a contributory role of PAF. First, the genome-wide occupancy of PAF (PAF1 and CDC73) is strikingly similar to the profiles of 3′-processing factors. Second, PAF depletion provokes widespread readthrough transcription. Third, PAF interacts with CPSF and CstF. Fourth, the recruitment of 3′-RNA processing factors to transcribing Pol II is impaired upon PAF depletion (Rozenblatt-Rosen et al., 2009). Moreover, our data also suggest

contributory roles of SPT5 and splicing factors, which appeared as core BRD4 interactors and have been implicated in 3′-RNA processing and termination (Kyburz et al., 2006; Davidson and West, 2013; Mayer et al., 2012; Cortazar et al., 2019; Parua et al., 2020).

This study identifies BRD4 as the BET protein that acts as a key regulator of Pol II transcription underlying the general 5′-elongation control point to ensure productive elongation, 3′-RNA processing, and transcription termination. Given the implication of BRD4 in human malignancies (Shi and Vakoc, 2014; Valent and Zuber, 2014), investigating how alterations of the link between the elongation checkpoint and 3′-RNA processing contribute to disease phenotypes will be an interesting future direction.

Limitations of the study

First, in this study, we have used immortalized human cell lines (K562 and MOLT4). Therefore, our conclusions cannot directly be transferred to other cell types or whole organisms. Second, our data suggest that the observed defects were partially mediated through BRD4 interactors including PAF and SPT5. To clarify their direct role, it requires specific perturbations of these factors in the same cell line and at a similar high kinetic resolution as applied in this study. Third, identified gene fractions are likely incomplete as NET-seq data generally suffers from low sequencing coverage due to stringent removal of artifacts that commonly arise from current library preparation methods (Gajos et al., 2021). Fourth, dTAG-BRD4 levels were lower than those of untagged BRD4. We cannot rule out that the reduced BRD4 levels caused cellular adaptations that may have contributed to less pronounced transcriptional defects observed upon BRD4-specific degradation. And fifth, based on our data, it remains unclear whether BRD4 physically binds to the identified interaction partners. Although our data suggest direct interactions, this needs further investigation.

STAR★METHODS

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

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Cell culture
- METHOD DETAILS
  - CRISPR/Cas9 genome editing
  - Degrader treatment
  - Immunoblotting
  - SI-NET-seq
  - Nascent RNA analysis by Oxford nanopore sequencing (nascONT-seq)
  - ChIP-Rx analysis
  - Sequencing data normalization
  - RLE with spike-ins
  - RNA-sequencing
  - Meta-gene analysis
  - 3′-RNA cleavage assay
  - Proteomics analysis
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.molcel.2021.06.026.

ACKNOWLEDGMENTS

We thank L.S. Churchman, L. Tora, and the Mayer lab for critical comments on the manuscript. We thank G. Winter and M. Jaeger for sharing dTAG7 and vectors for tagging. We thank S. Freier for cell culture support and T. Kratz for IP optimization. We thank B. Timmermann and the MPIMG Sequencing facility for sequencing. We thank the MS and fluorescence-activated cell sorting (FACS) facilities for help. This work was funded by the Max Planck Society (to A.M.) and the Deutsche Forschungsgemeinschaft (DFG, grant 418415292 to A.M. and the IRTG 2403 to A.M. and M.A.). O.J. was supported by a 2017 FEBS Long-Term Fellowship.

AUTHOR CONTRIBUTIONS

A.B., A.M., and M.A. developed SI-NET-seq. M.A. conducted experiments, analyzed proteomics data, and helped write the manuscript. A.B. established computational analysis pipelines, performed integrative data analysis, and helped write the manuscript. O.J. established and performed nascONT-seq. D.M. helped with MS data analysis. A.M. planned and designed experiments, wrote the manuscript, and had overall responsibility over the study.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: September 24, 2020
Revised: April 14, 2021
Accepted: June 22, 2021
Published: July 28, 2021

REFERENCES

Anamika, K., Gyenis, À., and Tora, L. (2013). How to stop: the mysterious links among RNA polymerase II occupancy 3′ of genes, mRNA 3′ processing and termination. Transcription *4*, 7–12.

Anand, P., Brown, J.D., Lin, C.Y., Qi, J., Zhang, R., Artero, P.C., Alaiti, M.A., Bullard, J., Alazem, K., Margulies, K.B., et al. (2013). BET bromodomains mediate transcriptional pause release in heart failure. Cell *154*, 569–582.

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

Baejen, C., Andreani, J., Torkler, P., Battaglia, S., Schwalb, B., Lidschreiber, M., Maier, K.C., Boltendahl, A., Rus, P., Esslinger, S., et al. (2017). Genome-wide Analysis of RNA Polymerase II Termination at Protein-Coding Genes. Mol. Cell *66*, 38–49.e6.

Baluapuri, A., Hofstetter, J., Dudvarski Stankovic, N., Endres, T., Bhandare, P., Vos, S.M., Adhikari, B., Schwarz, J.D., Narain, A., Vogt, M., et al. (2019). MYC Recruits SPT5 to RNA Polymerase II to Promote Processive Transcription Elongation. Mol. Cell *74*, 674–687.e11.

Belkina, A.C., and Denis, G.V. (2012). BET domain co-regulators in obesity, inflammation and cancer. Nat. Rev. Cancer *12*, 465–477.

Benson, D.A., Cavanaugh, M., Clark, K., Karsch-Mizrachi, I., Lipman, D.J., Ostell, J., and Sayers, E.W. (2013). GenBank. Nucleic Acids Res. *41*, D36–D42.

Bentley, D.L. (2014). Coupling mRNA processing with transcription in time and space. Nat. Rev. Genet. *15*, 163–175.

Brannan, K., Kim, H., Erickson, B., Glover-Cutter, K., Kim, S., Fong, N., Kiemele, L., Hansen, K., Davis, R., Lykke-Andersen, J., and Bentley, D.L. (2012). mRNA decapping factors and the exonuclease Xrn2 function in widespread premature termination of RNA polymerase II transcription. Mol. Cell *46*, 311–324.

Buratowski, S. (2009). Progression through the RNA polymerase II CTD cycle. Mol. Cell *36*, 541–546.

Chan, P.P., and Lowe, T.M. (2016). GtRNAdb 2.0: an expanded database of transfer RNA genes identified in complete and draft genomes. Nucleic Acids Res. *44* (D1), D184–D189.

Chan, S.L., Huppertz, I., Yao, C., Weng, L., Moresco, J.J., Yates, J.R., 3rd, Ule, J., Manley, J.L., and Shi, Y. (2014). CPSF30 and Wdr33 directly bind to AAUAAA in mammalian mRNA 3′ processing. Genes Dev. *28*, 2370–2380.

Chen, F.X., Xie, P., Collings, C.K., Cao, K., Aoi, Y., Marshall, S.A., Rendleman, E.J., Ugarenko, M., Ozark, P.A., Zhang, A., et al. (2017). PAF1 regulation of promoter-proximal pause release via enhancer activation. Science *357*, 1294–1298.

Chen, Y., Yamaguchi, Y., Tsugeno, Y., Yamamoto, J., Yamada, T., Nakamura, M., Hisatake, K., and Handa, H. (2009). DSIF, the Paf1 complex, and Tat-SF1 have nonredundant, cooperative roles in RNA polymerase II elongation. Genes Dev. *23*, 2765–2777.

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.

Chen, F.X., Smith, E.R., and Shilatifard, A. (2018). Born to run: control of transcription elongation by RNA polymerase II. Nat. Rev. Mol. Cell Biol. *19*, 464–478.

Cheung, K.L., Zhang, F., Jaganathan, A., Sharma, R., Zhang, Q., Konuma, T., Shen, T., Lee, J.-Y., Ren, C., Chen, C.-H., et al. (2017). Distinct Roles of Brd2 and Brd4 in Potentiating the Transcriptional Program for Th17 Cell Differentiation. Mol. Cell *65*, 1068–1080.e5.

Concordet, J.-P., and Haeussler, M. (2018). CRISPOR: intuitive guide selection for CRISPR/Cas9 genome editing experiments and screens. Nucleic Acids Res. *46* (W1), W242–W245.

Core, L., and Adelman, K. (2019). Promoter-proximal pausing of RNA polymerase II: a nexus of gene regulation. Genes Dev. *33*, 960–982.

Cortazar, M.A., Sheridan, R.M., Erickson, B., Fong, N., Glover-Cutter, K., Brannan, K., and Bentley, D.L. (2019). Control of RNA Pol II Speed by PNUTS-PP1 and Spt5 Dephosphorylation Facilitates Termination by a "Sitting Duck Torpedo" Mechanism. Mol. Cell *76*, 896–908.e4.

Cox, J., and Mann, M. (2008). MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. Nat. Biotechnol. *26*, 1367–1372.

Dantonel, J.C., Murthy, K.G., Manley, J.L., and Tora, L. (1997). Transcription factor TFIID recruits factor CPSF for formation of 3′ end of mRNA. Nature *389*, 399–402.

Davidson, L., and West, S. (2013). Splicing-coupled 3′ end formation requires a terminal splice acceptor site, but not intron excision. Nucleic Acids Res. *41*, 7101–7114.

Davidson, L., Muniz, L., and West, S. (2014). 3′ end formation of pre-mRNA and phosphorylation of Ser2 on the RNA polymerase II CTD are reciprocally coupled in human cells. Genes Dev. *28*, 342–356.

Davis, C.A., Hitz, B.C., Sloan, C.A., Chan, E.T., Davidson, J.M., Gabdank, I., Hilton, J.A., Jain, K., Baymuradov, U.K., Narayanan, A.K., et al. (2018). The Encyclopedia of DNA elements (ENCODE): data portal update. Nucleic Acids Res. *46*, D794–D801.

Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics *29*, 15–21.

Drexler, H.L., Choquet, K., and Churchman, L.S. (2020). Splicing Kinetics and Coordination Revealed by Direct Nascent RNA Sequencing through Nanopores. Mol. Cell *77*, 985–998.e8.

Eaton, J.D., and West, S. (2020). Termination of Transcription by RNA Polymerase II: BOOM! Trends Genet. *36*, 664–675.

Eaton, J.D., Davidson, L., Bauer, D.L.V., Natsume, T., Kanemaki, M.T., and West, S. (2018). Xrn2 accelerates termination by RNA polymerase II, which is underpinned by CPSF73 activity. Genes Dev. *32*, 127–139.

Eaton, J.D., Francis, L., Davidson, L., and West, S. (2020). A unified allosteric/torpedo mechanism for transcriptional termination on human protein-coding genes. Genes Dev. *34*, 132–145.

Ellison, M.A., Lederer, A.R., Warner, M.H., Mavrich, T.N., Raupach, E.A., Heisler, L.E., Nislow, C., Lee, M.T., and Arndt, K.M. (2019). The Paf1 Complex Broadly Impacts the Transcriptome of *Saccharomyces cerevisiae*. Genetics *212*, 711–728.

Endoh, M., Zhu, W., Hasegawa, J., Watanabe, H., Kim, D.-K., Aida, M., Inukai, N., Narita, T., Yamada, T., Furuya, A., et al. (2004). Human Spt6 stimulates transcription elongation by RNA polymerase II in vitro. Mol. Cell. Biol. *24*, 3324–3336.

Fitz, J., Neumann, T., Steininger, M., Wiedemann, E.-M., Garcia, A.C., Athanasiadis, A., Schoeberl, U.E., and Pavri, R. (2020). Spt5-mediated enhancer transcription directly couples enhancer activation with physical promoter interaction. Nat. Genet. *52*, 505–515.

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.

Frankish, A., Diekhans, M., Ferreira, A.-M., Johnson, R., Jungreis, I., Loveland, J., Mudge, J.M., Sisu, C., Wright, J., Armstrong, J., et al. (2019). GENCODE reference annotation for the human and mouse genomes. Nucleic Acids Res. *47* (D1), D766–D773.

Gajos, M., Jasnovidova, O., van Bömmel, A., Freier, S., Vingron, M., and Mayer, A. (2021). Conserved DNA sequence features underlie pervasive RNA polymerase pausing. Nucleic Acids Res. *49*, 4402–4420.

Glover-Cutter, K., Kim, S., Espinosa, J., and Bentley, D.L. (2008). RNA polymerase II pauses and associates with pre-mRNA processing factors at both ends of genes. Nat. Struct. Mol. Biol. *15*, 71–78.

Gregersen, L.H., Mitter, R., and Svejstrup, J.Q. (2020). Using TT_{chem}-seq for profiling nascent transcription and measuring transcript elongation. Nat. Protoc. *15*, 604–627.

Hou, L., Wang, Y., Liu, Y., Zhang, N., Shamovsky, I., Nudler, E., Tian, B., and Dynlacht, B.D. (2019). Paf1C regulates RNA polymerase II progression by modulating elongation rate. Proc. Natl. Acad. Sci. USA *116*, 14583–14592.

Hsu, S.C., Gilgenast, T.G., Bartman, C.R., Edwards, C.R., Stonestrom, A.J., Huang, P., Emerson, D.J., Evans, P., Werner, M.T., Keller, C.A., et al. (2017). The BET Protein BRD2 Cooperates with CTCF to Enforce Transcriptional and Architectural Boundaries. Mol. Cell *66*, 102–116.e7.

Jang, M.K., Mochizuki, K., Zhou, M., Jeong, H.-S., Brady, J.N., and Ozato, K. (2005). The bromodomain protein Brd4 is a positive regulatory component of P-TEFb and stimulates RNA polymerase II-dependent transcription. Mol. Cell *19*, 523–534.

Jonkers, I., Kwak, H., and Lis, J.T. (2014). Genome-wide dynamics of Pol II elongation and its interplay with promoter proximal pausing, chromatin, and exons. eLife *3*, e02407.

Kamieniarz-Gdula, K., Gdula, M.R., Panser, K., Nojima, T., Monks, J., Wiśniewski, J.R., Riepsaame, J., Brockdorff, N., Pauli, A., and Proudfoot, N.J. (2019). Selective Roles of Vertebrate PCF11 in Premature and Full-Length Transcript Termination. Mol. Cell *74*, 158–172.e9.

Kim, M., Krogan, N.J., Vasiljeva, L., Rando, O.J., Nedea, E., Greenblatt, J.F., and Buratowski, S. (2004). The yeast Rat1 exonuclease promotes transcription termination by RNA polymerase II. Nature *432*, 517–522.

Kim, J., Guermah, M., and Roeder, R.G. (2010). The human PAF1 complex acts in chromatin transcription elongation both independently and cooperatively with SII/TFIIS. Cell *140*, 491–503.

Kozomara, A., Birgaoanu, M., and Griffiths-Jones, S. (2019). miRBase: from microRNA sequences to function. Nucleic Acids Res. *47* (D1), D155–D162.

Krebs, A.R., Imanci, D., Hoerner, L., Gaidatzis, D., Burger, L., and Schübeler, D. (2017). Genome-wide Single-Molecule Footprinting Reveals High RNA Polymerase II Turnover at Paused Promoters. Mol. Cell *67*, 411–422.e4.

Kumar, A., Clerici, M., Muckenfuss, L.M., Passmore, L.A., and Jinek, M. (2019). Mechanistic insights into mRNA 3′-end processing. Curr. Opin. Struct. Biol. *59*, 143–150.

Kyburz, A., Friedlein, A., Langen, H., and Keller, W. (2006). Direct interactions between subunits of CPSF and the U2 snRNP contribute to the coupling of pre-mRNA 3′ end processing and splicing. Mol. Cell *23*, 195–205.

Laitem, C., Zaborowska, J., Isa, N.F., Kufs, J., Dienstbier, M., and Murphy, S. (2015). CDK9 inhibitors define elongation checkpoints at both ends of RNA polymerase II-transcribed genes. Nat. Struct. Mol. Biol. *22*, 396–403.

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

Li, H. (2018). Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics *34*, 3094–3100.

Li, B., and Dewey, C.N. (2011). RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics *12*, 323.

Li, X., Baek, G., Ramanand, S.G., Sharp, A., Gao, Y., Yuan, W., Welti, J., Rodrigues, D.N., Dolling, D., Figueiredo, I., et al. (2018). BRD4 Promotes DNA Repair and Mediates the Formation of TMPRSS2-ERG Gene Rearrangements in Prostate Cancer. Cell Rep. *22*, 796–808.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R.; 1000 Genome Project Data Processing Subgroup (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics *25*, 2078–2079.

Li, J., Xu, X., Tiwari, M., Chen, Y., Fuller, M., Bansal, V., Tamayo, P., Das, S., Ghosh, P., and Sen, G.L. (2021). SPT6 promotes epidermal differentiation and blockade of an intestinal-like phenotype through control of transcriptional elongation. Nat. Commun. *12*, 784.

Liao, Y., Smyth, G.K., and Shi, W. (2014). featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics *30*, 923–930.

Lindstrom, D.L., Squazzo, S.L., Muster, N., Burckin, T.A., Wachter, K.C., Emigh, C.A., McCleery, J.A., Yates, J.R., 3rd, and Hartzog, G.A. (2003). Dual roles for Spt5 in pre-mRNA processing and transcription elongation revealed by identification of Spt5-associated proteins. Mol. Cell. Biol. *23*, 1368–1378.

Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. *15*, 550.

Lovén, J., Hoke, H.A., Lin, C.Y., Lau, A., Orlando, D.A., Vakoc, C.R., Bradner, J.E., Lee, T.I., and Young, R.A. (2013). Selective inhibition of tumor oncogenes by disruption of super-enhancers. Cell *153*, 320–334.

Lykke-Andersen, S., Žumer, K., Molska, E.Ś., Rouvière, J.O., Wu, G., Demel, C., Schwalb, B., Schmid, M., Cramer, P., and Jensen, T.H. (2021). Integrator is a genome-wide attenuator of non-productive transcription. Mol. Cell *81*, 514–529.e6.

Mandel, C.R., Kaneko, S., Zhang, H., Gebauer, D., Vethantham, V., Manley, J.L., and Tong, L. (2006). Polyadenylation factor CPSF-73 is the pre-mRNA 3′-end-processing endonuclease. Nature *444*, 953–956.

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet. J. *17*, 10–12.

Mayer, A., and Churchman, L.S. (2016). Genome-wide profiling of RNA polymerase transcription at nucleotide resolution in human cells with native elongating transcript sequencing. Nat. Protoc. *11*, 813–833.

Mayer, A., Schreieck, A., Lidschreiber, M., Leike, K., Martin, D.E., and Cramer, P. (2012). The spt5 C-terminal region recruits yeast 3′ RNA cleavage factor I. Mol. Cell. Biol. *32*, 1321–1331.

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.

Mayer, A., Landry, H.M., and Churchman, L.S. (2017). Pause & go: from the discovery of RNA polymerase pausing to its functional implications. Curr. Opin. Cell Biol. *46*, 72–80.

Mi, H., Muruganujan, A., Ebert, D., Huang, X., and Thomas, P.D. (2019). PANTHER version 14: more genomes, a new PANTHER GO-slim and improvements in enrichment analysis tools. Nucleic Acids Res. *47* (D1), D419–D426.

Mohammed, H., Taylor, C., Brown, G.D., Papachristou, E.K., Carroll, J.S., and D'Santos, C.S. (2016). Rapid immunoprecipitation mass spectrometry of endogenous proteins (RIME) for analysis of chromatin complexes. Nat. Protoc. *11*, 316–326.

Muhar, M., Ebert, A., Neumann, T., Umkehrer, C., Jude, J., Wieshofer, C., Rescheneder, P., Lipp, J.J., Herzog, V.A., Reichholf, B., et al. (2018). SLAM-seq defines direct gene-regulatory functions of the BRD4-MYC axis. Science *360*, 800–805.

Nabet, B., Roberts, J.M., Buckley, D.L., Paulk, J., Dastjerdi, S., Yang, A., Leggett, A.L., Erb, M.A., Lawlor, M.A., Souza, A., et al. (2018). The dTAG system for immediate and target-specific protein degradation. Nat. Chem. Biol. *14*, 431–441.

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.

Noe Gonzalez, M., Blears, D., and Svejstrup, J.Q. (2021). Causes and consequences of RNA polymerase II stalling during transcript elongation. Nat. Rev. Mol. Cell Biol. *22*, 3–21.

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.

Nojima, T., Tellier, M., Foxwell, J., Ribeiro de Almeida, C., Tan-Wong, S.M., Dhir, S., Dujardin, G., Dhir, A., Murphy, S., and Proudfoot, N.J. (2018). Deregulated Expression of Mammalian lncRNA through Loss of SPT6 Induces R-Loop Formation, Replication Stress, and Cellular Senescence. Mol. Cell *72*, 970–984.e7.

Nordick, K., Hoffman, M.G., Betz, J.L., and Jaehning, J.A. (2008). Direct interactions between the Paf1 complex and a cleavage and polyadenylation factor are revealed by dissociation of Paf1 from RNA polymerase II. Eukaryot. Cell *7*, 1158–1167.

O'Leary, N.A., Wright, M.W., Brister, J.R., Ciufo, S., Haddad, D., McVeigh, R., Rajput, B., Robbertse, B., Smith-White, B., Ako-Adjei, D., et al. (2016). Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. Nucleic Acids Res. *44* (D1), D733–D745.

Ong, S.-E., and Mann, M. (2007). Stable isotope labeling by amino acids in cell culture for quantitative proteomics. Methods Mol. Biol. *359*, 37–52.

Orlando, D.A., Chen, M.W., Brown, V.E., Solanki, S., Choi, Y.J., Olson, E.R., Fritz, C.C., Bradner, J.E., and Guenther, M.G. (2014). Quantitative ChIP-Seq normalization reveals global modulation of the epigenome. Cell Rep. *9*, 1163–1170.

Parua, P.K., Kalan, S., Benjamin, B., Sansó, M., and Fisher, R.P. (2020). Distinct Cdk9-phosphatase switches act at the beginning and end of elongation by RNA polymerase II. Nat. Commun. *11*, 4338.

Peterlin, B.M., and Price, D.H. (2006). Controlling the elongation phase of transcription with P-TEFb. Mol. Cell *23*, 297–305.

Porrua, O., and Libri, D. (2015). Transcription termination and the control of the transcriptome: why, where and how to stop. Nat. Rev. Mol. Cell Biol. *16*, 190–202.

Potter, S.C., Luciani, A., Eddy, S.R., Park, Y., Lopez, R., and Finn, R.D. (2018). HMMER web server: 2018 update. Nucleic Acids Res. *46* (W1), W200–W204.

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

Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics *26*, 841–842.

Rahl, P.B., Lin, C.Y., Seila, A.C., Flynn, R.A., McCuine, S., Burge, C.B., Sharp, P.A., and Young, R.A. (2010). c-Myc regulates transcriptional pause release. Cell *141*, 432–445.

Rahman, S., Sowa, M.E., Ottinger, M., Smith, J.A., Shi, Y., Harper, J.W., and Howley, P.M. (2011). The Brd4 extraterminal domain confers transcription activation independent of pTEFb by recruiting multiple proteins, including NSD3. Mol. Cell. Biol. *31*, 2641–2652.

Ramírez, F., Dündar, F., Diehl, S., Grüning, B.A., and Manke, T. (2014). deepTools: a flexible platform for exploring deep-sequencing data. Nucleic Acids Res. *42*, W187–W191.

Richard, P., and Manley, J.L. (2009). Transcription termination by nuclear RNA polymerases. Genes Dev. *23*, 1247–1269.

Robinson, J., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E.S., Getz, G., and Mesirov, J.P. (2011). Integrative genomics viewer. Nat. Biotechnol. *29*, 24–26.

Rozenblatt-Rosen, O., Nagaike, T., Francis, J.M., Kaneko, S., Glatt, K.A., Hughes, C.M., LaFramboise, T., Manley, J.L., and Meyerson, M. (2009). The tumor suppressor Cdc73 functionally associates with CPSF and CstF 3′ mRNA processing factors. Proc. Natl. Acad. Sci. USA *106*, 755–760.

Sakuma, T., Nakade, S., Sakane, Y., Suzuki, K.T., and Yamamoto, T. (2016). MMEJ-assisted gene knock-in using TALENs and CRISPR-Cas9 with the PITCh systems. Nat. Protoc. *11*, 118–133.

Sansó, M., Levin, R.S., Lipp, J.J., Wang, V.Y.-F., Greifenberg, A.K., Quezada, E.M., Ali, A., Ghosh, A., Larochelle, S., Rana, T.M., et al. (2016). P-TEFb regulation of transcription termination factor Xrn2 revealed by a chemical genetic screen for Cdk9 substrates. Genes Dev. *30*, 117–131.

Schick, S., Grosche, S., Kohl, K.E., Drpic, D., Jaeger, M.G., Marella, N.C., Imrichova, H., Lin, J.G., Hofstätter, G., Schuster, M., et al. (2021). Acute BAF perturbation causes immediate changes in chromatin accessibility. Nat. Genet. *53*, 269–278.

Schröder, S., Cho, S., Zeng, L., Zhang, Q., Kaehlcke, K., Mak, L., Lau, J., Bisgrove, D., Schnölzer, M., Verdin, E., et al. (2012). Two-pronged binding with bromodomain-containing protein 4 liberates positive transcription elongation factor b from inactive ribonucleoprotein complexes. J. Biol. Chem. *287*, 1090–1099.

Schwalb, B., Michel, M., Zacher, B., Frühauf, K., Demel, C., Tresch, A., Gagneur, J., and Cramer, P. (2016). TT-seq maps the human transient transcriptome. Science *352*, 1225–1228.

Shi, J., and Vakoc, C.R. (2014). The mechanisms behind the therapeutic activity of BET bromodomain inhibition. Mol. Cell *54*, 728–736.

Sun, Y., Hamilton, K., and Tong, L. (2020). Recent molecular insights into canonical pre-mRNA 3′-end processing. Transcription *11*, 83–96.

Szklarczyk, D., Gable, A.L., Lyon, D., Junge, A., Wyder, S., Huerta-Cepas, J., Simonovic, M., Doncheva, N.T., Morris, J.H., Bork, P., et al. (2019). STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. Nucleic Acids Res. *47* (D1), D607–D613.

Tyanova, S., Temu, T., Sinitcyn, P., Carlson, A., Hein, M.Y., Geiger, T., Mann, M., and Cox, J. (2016). The Perseus computational platform for comprehensive analysis of (prote)omics data. Nat. Methods *13*, 731–740.

Untergasser, A., Cutcutache, I., Koressaar, T., Ye, J., Faircloth, B.C., Remm, M., and Rozen, S.G. (2012). Primer3–new capabilities and interfaces. Nucleic Acids Res. *40*, e115.

Uppal, S., Gegonne, A., Chen, Q., Thompson, P.S., Cheng, D., Mu, J., Meerzaman, D., Misra, H.S., and Singer, D.S. (2019). The Bromodomain Protein 4 Contributes to the Regulation of Alternative Splicing. Cell Rep. *29*, 2450–2460.e5.

Valent, P., and Zuber, J. (2014). BRD4: a BET(ter) target for the treatment of AML? Cell Cycle *13*, 689–690.

van de Peppel, J., Kemmeren, P., van Bakel, H., Radonjic, M., van Leenen, D., and Holstege, F.C.P. (2003). Monitoring global messenger RNA changes in externally controlled microarray experiments. EMBO Rep. *4*, 387–393.

Vos, S.M., Farnung, L., Boehning, M., Wigge, C., Linden, A., Urlaub, H., and Cramer, P. (2018a). Structure of activated transcription complex Pol II-DSIF-PAF-SPT6. Nature *560*, 607–612.

Vos, S.M., Farnung, L., Urlaub, H., and Cramer, P. (2018b). Structure of paused transcription complex Pol II-DSIF-NELF. Nature *560*, 601–606.

Vos, S.M., Farnung, L., Linden, A., Urlaub, H., and Cramer, P. (2020). Structure of complete Pol II-DSIF-PAF-SPT6 transcription complex reveals RTF1 allosteric activation. Nat. Struct. Mol. Biol. *27*, 668–677.

Wada, T., Takagi, T., Yamaguchi, Y., Ferdous, A., Imai, T., Hirose, S., Sugimoto, S., Yano, K., Hartzog, G.A., Winston, F., et al. (1998). DSIF, a novel transcription elongation factor that regulates RNA polymerase II processivity, is composed of human Spt4 and Spt5 homologs. Genes Dev. *12*, 343–356.

Wain, H.M., Lush, M., Ducluzeau, F., and Povey, S. (2002). Genew: the human gene nomenclature database. Nucleic Acids Res. *30*, 169–171.

Wang, R., Nambiar, R., Zheng, D., and Tian, B. (2018). PolyA_DB 3 catalogs cleavage and polyadenylation sites identified by deep sequencing in multiple genomes. Nucleic Acids Res. *46* (D1), D315–D319.

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.

Wheeler, T.J., and Eddy, S.R. (2013). nhmmer: DNA homology search with profile HMMs. Bioinformatics *29*, 2487–2489.

Winter, G.E., Mayer, A., Buckley, D.L., Erb, M.A., Roderick, J.E., Vittori, S., Reyes, J.M., di Iulio, J., Souza, A., Ott, C.J., et al. (2017). BET Bromodomain Proteins Function as Master Transcription Elongation Factors Independent of CDK9 Recruitment. Mol. Cell *67*, 5–18.e19.

Wissink, E.M., Vihervaara, A., Tippens, N.D., and Lis, J.T. (2019). Nascent RNA analyses: tracking transcription and its regulation. Nat. Rev. Genet. *20*, 705–723.

Yamaguchi, Y., Takagi, T., Wada, T., Yano, K., Furuya, A., Sugimoto, S., Hasegawa, J., and Handa, H. (1999). NELF, a multisubunit complex containing RD, cooperates with DSIF to repress RNA polymerase II elongation. Cell *97*, 41–51.

Yang, Z., Yik, J.H.N., Chen, R., He, N., Jang, M.K., Ozato, K., and Zhou, Q. (2005). Recruitment of P-TEFb for stimulation of transcriptional elongation by the bromodomain protein Brd4. Mol. Cell *19*, 535–545.

Yang, Y., Li, W., Hoque, M., Hou, L., Shen, S., Tian, B., and Dynlacht, B.D. (2016). PAF Complex Plays Novel Subunit-Specific Roles in Alternative Cleavage and Polyadenylation. PLoS Genet. *12*, e1005794.

Ye, J., Coulouris, G., Zaretskaya, I., Cutcutache, I., Rozen, S., and Madden, T.L. (2012). Primer-BLAST: a tool to design target-specific primers for polymerase chain reaction. BMC Bioinformatics *13*, 134.

Yu, M., Yang, W., Ni, T., Tang, Z., Nakadai, T., Zhu, J., and Roeder, R.G. (2015). RNA polymerase II-associated factor 1 regulates the release and phosphorylation of paused RNA polymerase II. Science *350*, 1383–1386.

Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers, R.M., Brown, M., Li, W., and Liu, X.S. (2008). Model-based analysis of ChIP-Seq (MACS). Genome Biol. *9*, R137.

Zhang, Q., Zeng, L., Shen, C., Ju, Y., Konuma, T., Zhao, C., Vakoc, C.R., and Zhou, M.-M. (2016). Structural Mechanism of Transcriptional Regulator NSD3 Recognition by the ET Domain of BRD4. Structure *24*, 1201–1208.

Zheng, B., Aoi, Y., Shah, A.P., Iwanaszko, M., Das, S., Rendleman, E.J., Zha, D., Khan, N., Smith, E.R., and Shilatifard, A. (2021). Acute perturbation strategies in interrogating RNA polymerase II elongation factor function in gene expression. Genes Dev. *35*, 273–285.

Zhou, Q., Li, T., and Price, D.H. (2012). RNA polymerase II elongation control. Annu. Rev. Biochem. *81*, 119–143.

Zorita, E., Cuscó, P., and Filion, G.J. (2015). Starcode: sequence clustering based on all-pairs search. Bioinformatics *31*, 1913–1919.

Please cite this article in press as: Arnold et al., A BRD4-mediated elongation control point primes transcribing RNA polymerase II for 3′-processing and termination, Molecular Cell (2021), https://doi.org/10.1016/j.molcel.2021.06.026

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
rabbit polyclonal anti-BRD4	Bethyl Labs	Cat#A301-985A50; RRID: AB_2631449
Rabbit polyclonal anti-CASP3	Cell Signaling Technology	Cat#9662; RRID: AB_331439
Rabbit polyclonal anti-CDC73	Bethyl Labs	Cat#A300-170A; RRID: AB_309449
Rabbit polyclonal anti-CPSF30	Bethyl Labs	Cat#301-584A; RRID: AB_1078872
Rabbit polyclonal anti-CPSF73	Bethyl Labs	Cat#A301-091A; RRID: AB_2084528
Rabbit polyclonal anti-CstF64	Bethyl Labs	Cat#A301-092A; RRID: AB_873014
Rabbit polyclonal anti-CstF77	Sigma	Cat#C0249; RRID: AB_10607433
Rabbit polyclonal anti-FIP1	Bethyl Labs	Cat#A301-461A; RRID: AB_999564
Mouse monoclonal anti-H2B	Santa Cruz Biotechnology	Cat#Sc-515808; RRID: N/A
Rabbit monoclonal anti-HA tag	Cell Signaling Technology	Cat#3724; RRID: AB_1549585
Rabbit polyclonal anti-PAF1	Abcam	Cat#Ab20662; RRID: AB_2159769
Rat monoclonal anti-RPB1 CTD Ser2-P	Active Motif	Cat#61083; RRID: AB_2687450
Rabbit polyclonal anti-RPB2	Genetex	Cat#TX102535; RRID: AB_1951313
Rabbit polyclonal anti-RTF1	Bethyl Labs	Cat#A300-179A; RRID:RRID: AB_2185963
Mouse monoclonal anti-SPT5	Santa Cruz Biotechnology	Cat#Sc-133217; RRID: AB_2196394
Rabbit polyclonal anti-SPT6	Novus Biologicals	Cat#NB100-2582; RRID: AB_2196402
Rabbit polyclonal anti-α-tubulin	abcam	Cat#ab18251; RRID: AB_2210057
Goat polyclonal anti-mouse IgG	LI-COR	Cat#926-32211; RRID: AB_621842
Goat polyclonal anti-rabbit IgG	LI-COR	Cat#926-32211; RRID: AB_621843
Chemicals, peptides, and recombinant proteins		
dBET6	MedChemExpress	Cat#HY-112588
dTAG7	Nabet et al., 2018/Georg Winter lab	N/A
Benzonase nuclease	EMD Millipore	Cat#E1014
Complete proteinase inhibitor cocktail	Roche	Cat#04693159001
PhosSTOP phosphatase inhibitor cocktail	Roche	Cat#04906845001
Superase.In	Invitrogen	Cat#AM2696
α-amanitin	Sigma-Aldrich	Cat#A2263
RNA fragmentation solution	NEB	Cat#E6150S
T4 RNA Ligase 2, truncated	NEB	Cat#M0242S
SuperScript III Reverse Transcriptase	Invitrogen	Cat#18080093
CircLigase ssDNA Ligase	Lucigen	Cat#CL4111K
Phusion HF DNA Polymerase	NEB	Cat#M0530S
TURBO DNase	Thermo Fisher Scientific	Cat#AM2238
E. coli poly(A) polymerase	NEB	Cat#M0276S
MTSEA biotin-XX linker	Biotinum	Cat#90066
Critical commercial assays		
NEBNext Ultra II DNA Library Prep Kit for Illumina	NEB	Cat#E7645S
NEBNext Multiplex Oligos for Illumina	NEB	Cat#E7335S
ProteoExtract Protein Precipitation kit	Merck Millipore	Cat#539180
QuantiTect Reverse Transcription kit	QIAGEN	Cat#205311
Direct cDNA Sequencing Kit	Oxford Nanopore Technologies	Cat#SQK-DCS109
R9.4.1 flow cell	Oxford Nanopore Technologies	Cat#FLO-MIN106D

(*Continued on next page*)

Please cite this article in press as: Arnold et al., A BRD4-mediated elongation control point primes transcribing RNA polymerase II for 3′-processing and termination, Molecular Cell (2021), https://doi.org/10.1016/j.molcel.2021.06.026

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Deposited data		
total RNA-seq MOLT4, DMSO 2 h, DMSO 6 h, dBET6 2 h, dBET6 6 h, JQ1 6 h	Winter et al., 2017	GEO: GSE79253
total RNA-seq K562	ENCODE Project Consortium	ENCODE: ENCSR109IQO (lab of Brenton Graveley)
RNA-seq NIH 3T3	ENCODE Project Consortium	ENCODE: ENCSR000CLW (lab of John Stamatoyannopoulos)
RNA-seq HeLa	ENCODE Project Consortium	ENCODE: ENCSR000CPR (lab of Thomas Gingeras)
RNA-seq mouse primary activated splenic B lymphocytes WT	Fitz et al., 2020	GEO: GSE132029
RNA-seq HCT116	Chen et al., 2017	GEO: GSE97527
nascONT-seq K562, DMSO 2 h, dBET6 2 h	This study	GEO: GSE158964
SI-NET-seq MOLT4, DMSO 2h, dBET6 2h	This study	GEO: GSE158963
SI-NET-seq K562 dTAG-BRD4, DMSO 2 h, dBET6 2 h, dTAG7 2 h, DMSO 40 min, DMSO 50 min, dBET6 40 min, dBET6 50 min	This study	GEO: GSE158963
NET-seq MOLT4, DMSO, JQ1	Winter et al., 2017	GEO: GSE79271
GRO-seq HCT116, shSCR, shPAF1	Chen et al., 2015	GEO: GSE70408
GRO-seq mouse primary activated splenic B lymphocytes, WT, SPT5 depletion	Fitz et al., 2020	GEO: GSE132029
ChIP-RX-seq MOLT4 for BRD4 and Ser2-P, 2h DMSO, 2h dBET6, 2h JQ1	Winter et al., 2017	GEO: GSE79288
ChIP-RX-seq K564 dTAG-BRD4 DMSO 2h dTAG7 2h for CPSF73, CstF64, FIP1, PAF1, Pol II Rpb2, Pol II Ser2-P, SPT5, Input	This study	GEO: GSE158965
ChIP-seq K562 BRD4	ENCODE Project Consortium	ENCODE: ENCSR583ACG (lab of Bradley Bernstein)
ChIP-seq K562 H3K36me3	ENCODE Project Consortium	ENCODE: ENCSR000AKR (lab of Bradley Bernstein)
ChIP-seq K562 H3K79me2	ENCODE Project Consortium	ENCODE: ENCSR000APD (lab of Bradley Bernstein)
ChIP-seq THP1 for LEO1, CDC73, PAF1, CTR9	Yu et al., 2015	GEO: GSE62171
ChIP-seq C2C12 for LEO1, CDC73, PAF1, CTR9, RTF1, WDR61	Yang et al., 2016	GEO: GSE72574
chromatin-MS K562 dTAG-BRD4, DMSO 2 h, dTAG7 2 h, see Table S3	This study	N/A
IP-MS K562 dTAG-BRD4, HA-BRD4, untreated, see Table S4	This study	N/A
total cell-MS, K562 dTAG-BRD4, dTAG7 2 h, DMSO 2 h, seeTable S1	This study	N/A
GENCODE Human v28, v29; Mouse M18, M22	Frankish et al., 2019	https://www.gencodegenes.org/
HUGO Gene Nomenclature	Wain et al., 2002	https://www.genenames.org/
polyA_DB v3.2	Wang et al., 2018	https://exon.apps.wistar.org/polya_db/v3/
RefSeq Release 109	O'Leary et al., 2016	https://www.ncbi.nlm.nih.gov/refseq/
miRBase v22.1	Kozomara et al., 2019	ftp://mirbase.org
UCSC's RepeatMasker Version 02/24/10		http://repeatmasker.org
GtRNAdb v2.0	Chan and Lowe, 2016	http://gtrnadb.ucsc.edu/
rDNA/RNA Mouse and Human	Benson et al., 2013	GenBank: U13369.1, U67616.1, X82564.1, GU372691.1

(*Continued on next page*)

Please cite this article in press as: Arnold et al., A BRD4-mediated elongation control point primes transcribing RNA polymerase II for 3′-processing and termination, Molecular Cell (2021), https://doi.org/10.1016/j.molcel.2021.06.026

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Experimental models: Cell lines		
MOLT4	ATCC	Cat#CRL-1582
K562	ATCC	Cat#CCL-243
K562 dTAG-BRD4	This study	N/A
NIH 3T3	ATCC	Cat#CRL-1658
Oligonucleotides		
primers for MBNL1, PVT1, CDC42, see Table S7	This study	N/A
sgPITCh: GCATCGTACGCGTACGTGTT	Sakuma et al., 2016	N/A
sgBRD4: ATGTCTGCGGAGAGCGGCCCTGG	Nabet et al., 2018/Georg Winter lab	N/A
tag-N-gF3: GACCTTACCCCTACGACGTG	This study	N/A
BRD4-F-new: CTTGGAGACCACAGCCAGAG	This study	N/A
BRD4-R-new: TTGAGCACCACTCTGAGCAG	This study	N/A
CMIP-F: GGCTGTGTTCAGACCATTCTTAGG	This study	N/A
CMIP-R: AGTCCTTAACCAGAATCTCAACCC	This study	N/A
Random decamer DNA linker: 5′-rApp/$(N)_{10}$CTGTAGGCACCATCAAT/3′-ddC	Gajos et al., 2021	N/A
random hexamer oligos	Invitrogen	Cat#N8080127
Recombinant DNA		
pX330A-sgPITCh-sgBRD4	Nabet et al., 2018/Georg Winter lab	N/A
pCRIS-PITChv2-dTAG-BRD4-puroR	Nabet et al., 2018/Georg Winter lab	N/A
pCRIS-PITChv2-dTAG-BRD4-blastiR	Nabet et al., 2018/Georg Winter lab	N/A
Software and algorithms		
RSEM v1.3.1	Li and Dewey, 2011	https://github.com/deweylab/RSEM
STAR 2.5.3a	Dobin et al., 2013	https://github.com/alexdobin/STAR
DEseq2 v1.25.4	Love et al., 2014	https://bioconductor.org/packages/release/bioc/html/DESeq2.html
Deeptools 2.0	Ramírez et al., 2014	https://deeptools.readthedocs.io/en/develop/
Bowtie2 v2.3.5.1	Langmead and Salzberg, 2012	http://bowtie-bio.sourceforge.net/bowtie2/index.shtml
IGV v2.8.2	Robinson et al., 2011	http://software.broadinstitute.org/software/igv/
HTSeq v0.11.4	Anders et al., 2015	https://htseq.readthedocs.io/en/master/
Guppy v3.2.4	Oxford Nanopore Technologies	https://community.nanoporetech.com/sso/login?next_url=%2F
Minimap2 v2.17	Li et al., 2018	https://github.com/lh3/minimap2
HMMER v3.3	Wheeler and Eddy, 2013	http://hmmer.org/
MACS2 v2.1.2	Zhang et al., 2008	https://pypi.org/project/MACS2/
Samtools v1.10	Li et al., 2009	http://samtools.sourceforge.net/
Bedtools v2.27.1	Quinlan and Hall, 2010	https://bedtools.readthedocs.io/en/latest/
PANTHER GO-slim	Mi et al., 2019	http://geneontology.org/
Subread v2.0.0	Liao et al., 2014	http://subread.sourceforge.net/
STRING	Szklarczyk et al., 2019	https://string-db.org/
PolyA_DB v3.2	Wang et al., 2018	https://exon.apps.wistar.org/PolyA_DB/

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
NCBI Primer BLAST	Ye et al., 2012	https://www.ncbi.nlm.nih.gov/tools/primer-blast/
Primer3	Untergasser et al., 2012	https://primer3.ut.ee/
CRISPOR v4.4	Concordet and Haeussler, 2018	http://crispor.tefor.net
MaxQuant v1.6.0.1	Cox and Mann, 2008	https://www.maxquant.org
Perseus	Tyanova et al., 2016	https://www.maxquant.org/perseus/
Cutadapt v2.4	http://journal.embnet.org/index.php/embnetjournal/article/view/200	https://github.com/marcelm/cutadapt
Starcode v1.1	Zorita et al., 2015	https://github.com/gui11aume/starcode

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Andreas Mayer (mayer@molgen.mpg.de).

Materials availability

The K562 dTAG-BRD4 cell line generated in this study is available on request with completion of MTA.

Data and code availability

All sequencing data (SI-NET-seq, nascONT-seq and ChIP-Rx) reported in this study are available at GEO: GSE158966.

Computational code that was generated and used in this study is available here:

SI-NET-seq: https://github.molgen.mpg.de/MayerGroup/net-seq-pipeline

nascONT-seq: https://github.molgen.mpg.de/MayerGroup/nascONT-seq-pipeline

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Cell culture

K562, K562 dTAG-BRD4 and MOLT4 cells were cultured in RPMI containing 10% FBS (Biochrom FBS Superior), 5% penicillin-streptomycin. NIH 3T3 cells were grown in DMEM containing 10% FBS (HyClone Cosmic Calf Serum, GE Healthcare; Bovine Calf Serum, iron-fortified, Sigma), 5% penicillin-streptomycin. Suspension cells were counted, centrifuged and re-seeded at 5×10^5 cells/ml every two days. NIH 3T3 were diluted to 2×10^6 cells/T75 flask every two days. Cells were kept in culture for not longer than four weeks. For quantitative proteomics experiments, K562 cells were cultured in the same way, but in the presence of heavy L-lysine and L-arginine and dialyzed FBS (see below.)

METHOD DETAILS

CRISPR/Cas9 genome editing

Design

Genome editing was done as described by Sakuma et al. (2016). For targeting the N terminus of BRD4 sgBRD4 (ATGTCTGCGG AGAGCGGCCCTGG) was expressed from pX330A-sgPITCh-sgBRD4, which also encoded the sgPITCh (GCATCGTACGC GTACGTGTT) for targeting the donor plasmid and SpCas9. The puroR/blastiR-P2A-2HA-FKBP12F36V tag, flanked by the 5′-microhomology arm (AGCATGTCTGCGGAGAGCGGC) and the 3′-microhomology arm (GGCCCTGGGACGAGATTGAGAA) were expressed from pCRIS-PITChv2-dTAG-BRD4-puroR or pCRIS-PITChv2-dTAG-BRD4-blastiR. CRISPOR (Concordet and Haeussler, 2018) was used to evaluate specificity and efficiency of the guide.

Cell line generation and validation

24 h prior to transfection, K562 cells were seeded at 2×10^5 cell/ml. Using the Nucleofector 2b device and the Amaxa Cell Line Nucleofector Kit V (Lonza), 1×10^6 cell were transfected with 2 μg of pX330A-sgPITCh-sgBRD4 and 1 μg of each pCRIS-PITChv2-dTAG-BRD4-puroR and pCRIS-PITChv2-dTAG-BRD4-blastiR. Selection with 0.5 μg/ml puromycin and 2.5 μg blasticidin was started after five days. To generate a monoclonal cell line, single cells were FACS-sorted into 96-well plates and expanded. Correct integration of the tag and homozygosity were confirmed using the PCR primers tag-N-gF3 (GACCTTACCCCTACGACGTG) and BRD4-R-new (TTGAGCACCACTCTGAGCAG), and BRD4-R-new and BRD4-F-new (CTTGGAGACCACAGCCAGAG), respectively. Presence

of unintended mutations flanking the integration site was excluded by amplicon sequencing. A possible off-target site was analyzed using CMIP-F (GGCTGTGTTCAGACCATTCTTAGG) and CMIP-R (AGTCCTTAACCAGAATCTCAACCC). Correct integration was additionally confirmed by western blotting.

To determine growth rates, K562 dTAG-BRD4 and unmodified K562 cells were seeded at 3 × 10^5 cells/ml in six biological replicates. Cell densities were determined after 24, 48 and 72 h using the automated counting system EVE (NanoEntek). Counting was done in duplicates. Doubling time was calculated using the formula:

$$T_{doubling} = T_{2-1}\left(\frac{ln2}{ln\frac{X_1}{X_2}}\right)$$

Degrader treatment

Cells reached their exponential growth phase after 48 h and were therefore treated at that time point. K562 cells and K562 dTAG-BRD4 cells at a concentration of 1 × 10^6/ml were treated with 100 nM dBET6 or 500 nM dTAG7. MOLT4 cells were treated with 100 nM dBET6. DMSO was used as a control. Treatment with each degrader at the given concentration resulted in a reduction of the targeted protein by > 90% after 2 h.

Immunoblotting

Washed pellets of a defined cell number were supplemented with 250–500 U benzonase and protease and phosphatase inhibitors and incubated on a shaker at 6°C for 15–30 min, followed by addition of 2X SDS buffer (100 mM Tris-HCl pH 7.0, 4% (wt/vol) SDS, 20% glycerol, 10% (vol/vol) β-mercaptoethanol) and heating to 95°C for 5 min. To obtain cytoplasmic, nucleoplasmic and chromatin samples, 1 × 10^7 cells were fractionated as described by Mayer and Churchman (2016). For SDS-PAGE, sample volumes were adjusted. Proteins were electrophoretically resolved on 4%–12% NuPAGE Bis-Tris or 8% Bolt Bis-Tris gels and visualized using a LI-COR Odyssey CLx imager.

SI-NET-seq

Spike-in procedure and cell fractionation

Spike-in native elongating transcript sequencing (SI-NET-seq) is based on the human NET-seq protocol by Mayer and Churchman (2016). 1.2 × 10^7 K562 dTAG-BRD4 cells per sample were treated with dBET6, dTAG7 or DMSO for 40, 50, 60 or 120 min. After treatment, cells were washed twice with ice-cold PBS and 2.4 × 10^6 mouse NIH 3T3 cells, which were used later for normalization, were added. Each sample consisting of K562 and NIH 3T3 cells at the ratio of 6:1 were subjected to cell fractionation and isolation of nascent RNA as described by Mayer and Churchman (2016). RNA was quantified using the Qubit HS RNA kit (Thermo Fisher Scientific).

SI-NET-seq library preparation

For library preparation, 3 μg RNA were used per sample. A DNA linker containing a random decamer sequence that served as a unique molecular identifier (UMI) was ligated to the 3′ end of the nascent RNA (Gajos et al., 2021). Ligated RNA was fragmented by incubating with NEBNext RNA fragmentation solution at 95°C for 10 min and separated on a 15% TBE/urea gel. Fragments between 55 and 140 nt were excised. cDNA was synthesized using SuperScriptIII according to the manufacturer's instructions. The obtained cDNA was separated on a 10% TBE/urea gel and fragments between 95 to 180 nt were excised. The following steps were performed as described by Mayer and Churchman (2016). SI-NET-seq libraries were sequenced in SR75 mode on a HiSeq 4000 or in SR100 mode on a NovaSeq 6000.

Masked genomic regions

We masked regions of the human and mouse reference genomes that are transcribed by Pol I and Pol III, as well as loci of short chromatin-associated RNAs. This includes: 5S, 7SK, HY1, HY3, HY4, HY5, LSU-rRNA_Hsa, RNase_MRP_RNA, RNase_P_RNA, SSU-rRNA_Hsa, U1, U2, U3, U4, U5, U6, U7, U8, U13, U14, U17, Y_RNA, antisense_RNA, guide_RNA, miRNA, misc_RNA, rRNA, rRNA_pseudogene, sRNA, scRNA, snRNA, snoRNA, tRNA, telomerase_RNA and vaultRNA. We used the annotations of GENCODE (human: v28 and v29; mouse: M18 and M22), RefSeq release 109 (O'Leary et al., 2016), miRBase v22.1 (Kozomara et al., 2019) and the UCSC's RepeatMasker annotation.

SI-NET-seq data processing

Processing of SI-NET-seq data was performed as previously described (Mayer et al., 2015; Gajos et al., 2021) with some modifications. Briefly, we trimmed adaptor sequences using cutadapt (v2.4) (Martin, 2011) with the following parameters: -a ATCTCG TATGCCGTCTTCTGCTTG -a AAAAAAAAAAGGGGGGGGGGGGGG -a GGGGGGGGGGGGGGGGGGGGGGGG -m 11 -e 0.2. To remove PCR duplicates, identical reads with the same unique molecular identifier (UMI) sequence were collapsed to one read using Starcode (v1.1) (Zorita et al., 2015). The ten 5′ end nucleotides corresponding to the UMI sequence were trimmed, but the sequence information remained associated with the read using an adapted custom Python script from (Mayer et al., 2015). The obtained reads

were aligned to the joined reference genome which we assembled from human (GRCh38.p12) and mouse (GRCm38.p6) using the STAR aligner (v2.5.3a) (Dobin et al., 2013) with the following parameters: -clip3pAdapterSeq ATCTCGTATGCCGTCTTCTGCTTG −3clip3pAdapterMMp 0.21 -clip3pAfterAdapterNbases 1 -outFilterMultimapNmax 1000000000 -outSJfilterOverhangMin 3 1 1 1 -outSJfilterDistToOtherSJmin 0 0 0 0 -alignIntronMin 11 -alignEndsType EndToEnd. We included additional rDNA sequences (GenBank: U13369.1, U67616.1, X82564.1, GU372691.1), and the processed tRNA from human and mouse (GtRNAdb v2.0) (Chan and Lowe, 2016). For uniquely mapped reads, the position corresponding to the 3′ end of the nascent RNA fragment was recorded with a custom Python script using the HTSeq package (Anders et al., 2015). Reads that originated from reverse transcription mispriming were identified as reads for which the UMI sequence was identical with the genomic sequence adjacent to the aligned read and were excluded. Reads that aligned to the same genomic position and contained an identical UMI sequence were considered as PCR duplicates and were removed. The 3′-most nucleotide position of introns and exons were masked to avoid any impact from splicing intermediates as described in detail previously (Mayer et al., 2015). Finally, we also masked regions as described in the previous section (*Masked genomic regions*).

Differential Pol II occupancy analysis

We statistically tested for changes in the Pol II occupancy using DEseq2 v1.25.4 (Love et al., 2014). First, we quantified the Pol II occupancy at the promoter-proximal regions (TSS to TSS +300 bp) and the gene-body regions (TSS +300 bp to pA site). We considered active non-overlapping genes with a minimum gene length of 1 kb which had at least six reads that mapped to the respective region. Second, we tested for significant changes in the regions of interest between the conditions using DEseq2. For data normalization, we used the RLE normalization with and without spike-ins (see *Relative Log Expression (RLE)* and *RLE with spike-ins*). Changes were considered as significant when the FDR adjusted p value was < 0.05.

BET-sensitive and BET-resistant gene definition

We defined BET-sensitive and BET-resistant genes according to their response upon BET protein degradation. BET-sensitive genes showed a significant reduction of Pol II occupancy over their gene-body region (TSS +300 bp to pA site, FDR adjusted p value < 0.05). On the contrary, BET-resistant genes showed no significant changes in Pol II occupancy over the gene-body region. We excluded genes from this analysis that had less than 6 reads mapping to their promoter-proximal region.

Transcriptional Readthrough Index (RTI) calculation

To compare the efficiency of Pol II transcription termination at a given gene *j* in different conditions, e.g., control versus treatment, we developed and defined the transcriptional 'Readthrough Index' (RTI) as follows:

$$RTI = ATD_j(Occ_{treatment}) - ATD_j(Occ_{control})$$

The RTI of a given gene quantifies the shift of the Pol II occupancy, termed the 'average transcription termination distance' (ATD), to the pA site between two conditions. The ATD summarizes the distribution of Pol II occupancy in the extended termination zone with the length *l* as the weighted distance to the pA site. We defined the ATD as:

$$ATD_j = \frac{\sum_{i=1}^{l} d_{pA_j}(i) \cdot Occ(i)}{\sum_{i=1}^{l} Occ(i)}$$

for $i = 1 \ldots l$ positions in the extended termination zone, where *Occ(i)* reports the Pol II occupancy at observation *i* and d_{pA_j} the distance to the last active pA site of gene *j*.

The extended termination zone of a gene *j* describes the variable region downstream of the last active pA site where we assumed transcription termination of Pol II to occur. Starting from the last active pA site, we defined the end of the extended termination zone where the Pol II signal dropped below the threshold *t* for the window size *W* in the control experiment and upon treatment. We calculated the RTI for SI-NET-seq, standard NET-seq and GRO-seq data with technology adjusted parameter settings. GRO-seq data were re-analyzed for available data upon PAF1 (Chen et al., 2015) and SPT5 (Fitz et al., 2020) knockdown. For SI-NET-seq and standard NET-seq, we used a window size *W* of 5 kb and a minimum signal strength *t* of 1. For the datasets with higher background noise levels (GRO-seq), we adjusted the minimum signal strength *t* to 10 or 100. By definition, the termination zone ends 1.5 kb upstream of the next active TSS to avoid spillover effects.

For the calculation of the RTI we considered only non-overlapping genes (TSS to pA +5 kb). The length of the termination zone of each condition had to be > 1 kb with an RPKM > 0.01 or > 0.1 for SI-NET-seq or GRO-seq, respectively. A gene with an RTI of $\geq$ 5 kb was defined as a readthrough gene. To determine the RTI for antisense transcription that usually accompanies sense transcription in the opposite orientation we calculated the ATD_{AS}, termed based on the Pol II occupancy signal from the opposite strand (antisense). The main difference to the calculation of the standard ATD was that ATD_{AS} uses the distance to the TSS instead of to the pA site. For ATD_{AS} calculation we used a window size *W* of 1 kb for non-overlapping genes (TSS −5 kb to pA).

Nascent RNA analysis by Oxford nanopore sequencing (nascONT-seq)

Treatment, cell fractionation and library preparation

1×10^8 K562 cells were treated with 100 nM dBET6 or an equivalent volume of DMSO for 1 h 50 min, and for another 10 min in the presence of 500 μM 4-thiouridine (4sU). All steps, which were adapted from Mayer and Churchman (2016), were performed at 4°C or

on ice, if not stated otherwise. K562 cells were transferred into two 50 mL falcon tubes and spun 2 min at 1,650 x g. After removal of medium cells were resuspended in 1 mL of lysis buffer (1X PBS, 0.15% (vol/vol) NP-40, 25 mM α-amanitin, protease inhibitors, 50 U/ml SUPERaseIN), incubated on ice for 2 min, and nuclei were collected by centrifugation for 3 min at 500 x g. Nuclei pellets were washed with 3 mL nuclei wash buffer (Mayer and Churchman, 2016) and centrifuged for 3 min at 500 x g. Next, nuclei were gently resuspended in 750 μl glycerol buffer and 750 μl nuclei lysis buffer were added. The mixture was pulse-vortexed five times for 5 min, incubated on ice for 2 min and spun at 18,400 x g. Chromatin pellets were dissolved in 1.5 mL Qiazol reagent, 1 mM DTT, 5 mM EDTA for 1 h at 40°C and 1,000 rpm. For RNA extraction, two chromatin pellets of the same treatment were combined. After addition of 0.2 volumes of chloroform, the RNA-containing phase was separated using MaXtract High Density 15 mL tubes (QIAGEN) according to the manufacturer's instructions. RNA was precipitated with 100% isopropanol on ice for 10 min followed by centrifugation for 20 min at 20,000 x g and 4°C. Next, the pellet was washed twice with 80% ice-cold ethanol and resuspended in nuclease-free H_2O. The samples were treated with TURBO DNase (Thermo Fisher Scientific) according to the manufacturer's instructions. The reaction was stopped by adding EDTA to the final concentration of 15 mM. RNA was purified by adding one volume of phenol:chloroform:isoamyl alcohol (25:24:1). The aqueous phase was separated using 5PRIME Phase Lock Gel Heavy 2 mL tubes by centrifugation for 5 min at 12,000 x g. The RNA was collected by isopropanol precipitation, washed with 85% ethanol and resuspended in nuclease-free H_2O.

The RNA was biotinylated using MTSEA biotin-XX linker (Biotinum) and purified by μMACS streptavidin MicroBeads as described by Gregersen et al. (2020) with the following modification: The column was washed three times with 1 mL of 65°C pre-warmed pull-out wash buffer, followed by three washes with 1 mL pull-out wash buffer at room temperature. Following elution, the RNA was purified using the ZYMO Clean & Concentrator-5 kit. Next, the nascent RNA was polyA tailed using *E. coli* polyA polymerase (Takara) for 7 min at 37°C. The reaction was stopped by adding EDTA. The RNA was purified using the Clean & Concentrator-5 kit (Zymo Research).

The library was prepared using the direct cDNA sequencing (SQK-DCS109) kit (Oxford Nanopore Technologies) according to the manufacturer's instructions with the following modification: To capture longer transcripts, reverse transcription and second-strand synthesis was extended to 2 h and 45 min, respectively. Sequencing was performed on a MinION sequencer for 48 to 72 h with −180 mV starting voltage using R9.4.1 FLO-MIN106 flow cells.

nascONT-seq data processing

For base calling we applied Guppy v3.2.4 using the kit and flow cell specific parameters '–flowcell FLO-MIN106–kit SQK-DCS109'. Next, we aligned the reads to the human reference genome GRCh38.p12 using minimap2 v2.17 (Li, 2018) with the parameters '-ax splice -ub -k14–secondary = no -O 12,32–junc-bonus = 19'. We improved the splice site detection using a reference splice junction set based on the GENCODE v28 gene annotation. The strand information was derived from Minimap2 for spliced reads. The unspliced reads remained unclassified. We quantified the transcript abundance of active genes (see *Annotation of active genes and gene types*) using featureCounts from the subreads package v2.0.0 (Liao et al., 2014) in the long-read ('-L') mode. Furthermore, we recognized a strong enrichment bias toward polyA- or polyT-rich sequence regions. For comparisons of different genomic regions, we removed the reads which mapped with their 3′ ends directly to stretches of polyA or polyT in the reference genome. Finally, we masked genomic regions as previously described in *Masked regions*.

Identification of full-length transcripts from nascONT-seq data

The direct cDNA sequencing (SQK-DCS109) kit by Oxford Nanopore Technologies introduced two primers, VPN and SSP, at the 3′- and 5′ ends of an RNA molecule, respectively. The valid full-length transcript was defined as a region mapped to the reference genome that is flanked by the VNP and SSP primers in proper orientation. To identify the primers, we built two Hidden Markov Models based on the primer sequences and iteratively improved them on our data. We used the 'nhmmscan' function of the HMMER v3.3 (Potter et al., 2018) software to identify sequences similar to the primers (E-value < 0.1) and added those to the Hidden Markov Models (two iterations). The optimized Hidden Markov models with the parameters '–max -E 10' were applied for the final primer search (VNP and SSP) in each sample. Most reads were not full-length transcripts but showed missing primers at the 5′- or 3′ ends (5′- or 3′-truncated). Furthermore, we observed reads lacking primer sequences or fused reads (primary and supplementary alignments).

We derived the type of a read (full-length transcripts, 3′-truncated, 5′-truncated, no primer, fused read) from:

- the identified primer types (VNP, SSP),
- the primer positions relative to the mapping region in the read, and the mapping strand of the primer alignments (+, -).

These features of a read *r* were encoded in an ordered sequence

$$P_r = \{s_{r0}, s_{r1}, \ldots, s_m\}$$

where

$$s \in \{Start, VNP^+, VNP^-, SSP^+, SSP^-, Primary, Supplementary, End\}$$

We identified for each read *r* the vector

$$V_r = v_{r1}, v_{r2}, \ldots, v_m$$

with $v_{ri} \in \{0, 1\}$ that maximized a score calculated by the function $S(P_r[V_r])$. The score S was calculated with the function:

$$S(P) = \prod_{i=0}^{n-1} T(s_i,\ s_{i+1}) \left(\sum_{i=0}^{n} w(s_i) f(s_i) - \sum_{i=0}^{n-1} 0.1 \cdot d(s_i, s_{i+1}) \right)$$

where

$$T = \begin{pmatrix} Start & VNP^{+} & VNP^{-} & SSP^{+} & SSP^{-} & Primary & Supplementary & End & \\ 0 & 1 & 0 & 1 & 0 & 1 & 1 & 0 & Start \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & VNP^{+} \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & VNP^{-} \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & SSP^{+} \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & SSP^{-} \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & Primary \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & Supplementary \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & End \end{pmatrix}$$

$$w(s) = \begin{cases} 0.5, & |if\ s \in Primary,\ Supplementary \\ 0.4, & |otherwise \end{cases}$$

$$f(s) = \begin{cases} 0, & |if\ s \in Start, End \\ Alignment\ score, & |otherwise \end{cases}$$

and with $d(s_i, s_{i+1})$ as a function of two states, reporting the distance between both states in percent of the read length. We derived the *Alignment score* from the result table of minimap2 and HMMER. Finally, we use $P_r[V_r]$ to define the read type as described in Table S6.

Calculation of pA site cleavage efficiency

To determine if the cleavage efficiency was significantly changed upon dBET6 treatment, we first identified actively used pA sites of active non-overlapping genes (+/−10 kb). For this purpose we used the human polyA_DB v3.2 (Wang et al., 2018) to quantify the number of 3′ ends of transcripts that mapped precisely to the pA site in the control (DMSO) experiment. For each gene, we selected the pA sites that had the highest coverage and the highest polyA_DB score. Next, we compared the ratio of transcripts mapping to the pA site (+/−20 bp) versus transcripts spanning the region between the conditions.

We considered all read types (see *Identification of full-length transcripts from nascONT-seq data*) as well as reads without strand classification. The results for each pA site were summarized in a contingency table where we tested for significant changes in cleavage efficiency using the one-sided Fisher's exact test ($p < 0.05$). Biological replicates were pooled together.

Transcript length analysis

We performed the read length analysis on the full-length transcripts identified as described in *Identification of full-length transcripts from nascONT-seq data*. The read length corresponds to the region of the read mapping to the reference genome.

ChIP-Rx analysis

Treatments, spike-in procedure and DNA purification

ChIP was performed as described by Baluapuri et al. (2019) with the following modifications: Instead of chromatin mouse cells were used as spike-ins. For each antibody and condition, two biological replicates were prepared and sequenced. 4×10^7 K562 dTAG-BRD4 cells were treated with dTAG7 or DMSO for 2 h, as described above, washed with PBS and crosslinked with 1% methanol-free formaldehyde for 5 min at room temperature. Formaldehyde crosslinking was stopped with 250 mM glycine in PBS for 5 min. Mouse NIH 3T3 cells were crosslinked for 8 min and rinsed with 250 mM glycine. K562 and NIH 3T3 cells were washed with ice-cold PBS and then combined at a ratio of 5:1. After cell lysis as described by Baluapuri et al. (2019), each sample was into two AFA fiber 1 mL tubes (Covaris) and sheared by sonication using the E220evolution sonicator (Covaris) for 20 min at 4°C using the following settings: intensity 4, duty cycle 5%, 200 cycles per burst. After pre-clearing, 2.5% of the sample volume was saved as input. For each sample 5–12 μg of antibody were incubated with 80 μl Dynabeads protein G for 1.5 h at room temperature on a rotating wheel, followed by an incubation for > 5 h at 4°C. Bead-coupled antibodies were added to the lysates and incubated at 4°C overnight on a rotating wheel. On the next day, beads were washed three times for 45 s

using the washing buffers described by Baluapuri et al. (2019). DNA fragments were eluted twice in elution buffer (100 mM $NaHCO_3$, 1% (wt/vol) SDS) at room temperature. ChIP and input samples were incubated for 1 h at 37°C with 4 μg RNase A and for 2 h at 50°C with 8 μg Proteinase K, before crosslinking was reversed at 65°C for 5 h in the presence of 0.84 M NaCl. DNA was purified using the ChIP DNA Clean & Concentrator kit (Zymo Research) and quantified using the Qubit dsDNA HS kit (Thermo Fisher Scientific).

ChIP-Rx library preparation

ChIP-Rx libraries were constructed from 6–30 ng of ChIP or input DNA using the NEBNext Ultra II DNA kit (NEB) according to the manufacturer's instructions. After purification of the amplified library with one volume AMPure beads (Beckman Coulter), fragments of 200 to 500 bp were size-selected from an 8% TBE gel (Thermo Fisher Scientific), as described for the SI-NET-seq library preparation. Samples were sequenced on an Illumina HiSeq 4000 or NovaSeq 6000 sequencer in PE75 or PE100 mode, respectively.

ChIP-Rx data processing

We aligned all reads to their respective reference genome (human/GRCh38.p12 or mouse/GRCm38.p6) using Bowtie2 v2.3.5.1 (Langmead and Salzberg, 2012) in single-end or paired-end mode with '-k 1'. We extracted the raw density for the protein of interest using bedtools (Quinlan and Hall, 2010) in the 'genomecov -bg' mode. For data normalization, we calculated the RPM values for between sample comparisons (see *Reads per Million (RPM)*) or RPK values for comparisons between genes within one sample (see *Reads per Kilobase (RPK)*).

For ChIP-Rx, we aligned the sequencing reads to a joined reference genome which we assembled from human (GRCh38.p12) and mouse (GRCm38.p6) reference genomes. Following read mapping, we separated the human and mouse tracks to calculate the RRPM values for quantitative comparisons (see *Reference adjusted Reads per Million (RRPM)*).

Sequencing data normalization

We used standard normalization techniques to remove effects originating from different sequencing depths, gene lengths, or both. The main limitation of current standard normalization strategies is the assumption that the majority of observations are not changed between conditions. To avoid misinterpretation in cases where global changes were expected, we additionally performed normalizations on spike-in controls from mouse cells or synthetic ERCC spike-in mixes.

Reads per Million (RPM)

'Reads per Million' (RPM) values were corrected for different sequencing depths. The normalization factor α was defined as:

$$\alpha = \frac{1}{N}$$

where N represents the total number of reads (in million) mapping to the reference genome. The RPM values are defined as:

$$RPM = c \cdot \alpha$$

where c represents the raw read counts. We applied this strategy for ChIP-Rx, RNA-seq and nascONT-seq data normalization.

Reads per Kilobase (RPK)

'Reads per Kilobase' (RPK) values were corrected for different gene lengths. The normalization vector α_j is defined as:

$$\alpha_j = \frac{1}{L_j}$$

for $j = 1 \ldots k$, where k is the total number of genes and L_j the length of the gene j (in kilobases). The RPK value for a gene j is defined as:

$$RPK_j = c_j \cdot \alpha_j$$

where c_j represents the raw read count of a given gene j. We used this normalization strategy to compare the signal of different genes within one sample of ChIP-Rx, SI-NET-seq, RNA-seq or nascONT-seq data.

Transcripts per Kilobase Million (TPM)

'Transcript per Kilobase Million' (TPM) values were normalized for different sequencing depths and gene lengths. First, we calculated the RPK values for each gene (see *Reads per Kilobase (RPK)*). Next, we computed the normalization factor α, which is defined as:

$$\alpha = \frac{1}{\sum_{j=1}^{k} RPK_j \cdot \frac{1}{1{,}000{,}000}}$$

for $j = 1 \ldots k$, where k is the total number of genes and RPK_j the RPK value of gene j. The TPM value for gene j is defined as:

$$TPM_j = RPK_j \cdot \alpha$$

We applied this normalization strategy to compare gene expression levels of RNA-seq and SI-NET-seq datasets.

Relative Log Expression (RLE)

For 'Relative Log Expression' normalization we used DESeq2 (Love et al., 2014). We used this normalization strategy for differential Pol II occupancy analysis using SI-NET-seq data.

Please cite this article in press as: Arnold et al., A BRD4-mediated elongation control point primes transcribing RNA polymerase II for 3′-processing and termination, Molecular Cell (2021), https://doi.org/10.1016/j.molcel.2021.06.026

RLE with spike-ins

The default configuration of DESeq2 assumes that the majority of observations does not change between conditions. To capture uniform changes with RLE, we applied spiked-in controls from mouse cells or a synthetic ERCC spike-in mix. We added the expression values with the 'controlGenes' parameter to calculate the 'size factor' exclusively on the spiked-in datasets.

We used this normalization strategy for the differential gene expression analysis using RNA-seq data and for the differential Pol II occupancy analysis using SI-NET-seq data.

Reference-adjusted Reads Per Million (RRPM)

'Reference-adjusted Reads Per Million' (RRPM) were calculated as described by Orlando et al. (2014). The normalization factor α is defined as:

$$\alpha = \frac{1}{N_{spike-in}}$$

where $N_{spike-in}$ represents the total number of reads (in million) mapping to the spiked-in reference genome. The RRPM values are defined as:

$$RRPM = c \cdot \alpha$$

where c represents the raw read counts. We used this normalization strategy for quantitative comparisons between SI-NET-seq profiles.

Pol II-based ChIP-Rx normalization

For Pol II-associated factors, we performed Pol II-based normalizations where we measured the ChIP-Rx signal in proportion to the Pol II ChIP-Rx signal. Instead of providing an input dataset for data normalization, we used total Pol II ChIP-seq signal. We applied MACS2 v2.1.2 (Zhang et al., 2008) in 'bdgcmp' mode with the parameters '-p 1 -m logFE'.

RNA-sequencing

Annotation of active genes and gene types

For gene and isoform quantification, we used RSEM v1.3.1 (Li and Dewey, 2011) in the single-end or paired-end mode with the parameters '–star–calc-pme', which uses the STAR v2.5.3a mapper (Dobin et al., 2013). We considered genes as actively transcribed where we observed a total gene expression with a TPM $\geq$ 1 or steady-state RNA expression with a TPM > 10 (see *Transcripts per Kilobase Million (TPM)*). All our analyses were done on the human (GRCh38.p12) or mouse reference genome (GRCm38.p6) using the annotated transcriptome from GENCODE v28 and GENCODE M18 (Frankish et al., 2019) respectively. Next, we refined the annotation of the active gene set by identifying the first active TSS and the last active pA site. We defined the first active TSS as the most upstream TSS from the set of transcripts of a given gene accounting for at least 10% of its gene activity. The last active pA site was defined similarly, with the exception of choosing the most downstream pA site.

Furthermore we used the GENCODE v28 biotype classifications 'protein-coding', 'miRNA', 'snRNA' and 'snoRNA' to define gene classes. For long non-coding RNAs (lncRNAs), we merged the genes listed as biotype 'lincRNA' and 'antisense'. Histone genes are defined as a subset of protein-coding genes where we used the HUGO Gene Nomenclature (HUGO: 864) (Wain et al., 2002).

Differential gene expression analysis with ERCC RNA spike-in mix

We aligned all RNA-seq data to the combined reference genome of human (GRCh38.p12) and ERCC spike-ins (ERCC92) using STAR v2.5.3a with default parameters in single-end mode. Next, we quantified the transcript abundance of the ERCC spike-ins and the annotated genes from GENCODE v28 using HTSeq v0.11.4 (Anders et al., 2015) in 'union' mode. For the differential gene expression analysis we tested for significant changes between the conditions using DESeq2 v1.25.4. For data normalization, we applied the RLE normalization considering the ERCC spike-ins (see *RLE with spike-ins*). We defined genes as significantly changed that had an FDR adjusted p value of < 0.05.

Classification of differentially expressed genes as 'readthrough-associated' or '-independent'

We defined 'readthrough-associated' and 'independent' genes to test the potential impact of readthrough transcription on the transcriptional output of neighboring gene units. We classified all differentially expressed genes into one of the two categories. A gene was considered to be 'readthrough-associated' if:

- the gene expression was significantly increased upon treatment, and the gene overlapped (+/−100 bp) with the extended termination zone of a readthrough gene on the same DNA strand, or
- the gene expression was significantly decreased upon treatment, and the gene overlapped (+/−100 bp) with the extended termination zone of a readthrough gene on the opposite DNA strand.

The remaining differentially expressed genes are readthrough 'independent' genes. The readthrough genes were defined based on SI-NET-seq data (2 h dBET6, see *Transcriptional Readthrough Index (RTI) calculation*) and total RNA-seq (2 h and 6 h dBET6) in MOLT4 cells. For total RNA-seq, we included genes with an increased transcript coverage (> 0.01 RPKM) in the termination zone compared to the control.

Splicing analysis

We determined the fraction of spliced RNA molecules per gene to identify the level of post-transcriptional processing from previous RNA-seq data (Winter et al., 2017). First, reads were mapped to the reference genome (see Differential gene expression analysis with ERCC RNA spike-in mix). Second, all reads of a gene were considered that span at least one intron-exon or exon-intron boundary. Next, remaining reads were classified as 'spliced' or 'not spliced'. A 'spliced' read maps to a non-contiguous loci of the corresponding gene. Reads mapping to only one contiguous locus are classified as 'not spliced'. The fraction of spliced RNA molecules per gene corresponds to the fraction of reads classified as 'spliced' reads. In the corresponding analysis we calculated this fraction for all actively transcribed genes in the control experiments and readthrough activated genes upon treatment. A gene is readthrough activated if:

- the TPM was < 1 in the control experiment, and
- the gene was classified as 'readthrough-associated' (see Classification of differentially expressed genes as readthrough-associated or independent genes)

Meta-gene analysis

For nascONT-seq, ChIP-Rx, ChIP-seq and SI-NET-seq data average profiles were calculated using Deeptools 2.0 (Ramírez et al., 2014) in the 'reference-point' mode. Furthermore, we used the 'scaled-region' option to adjust genes of different length from the promoter-proximal region (TSS +1 kb) to the pA site or for the entire gene region. Further, we customized the tool to make it applicable for strand-specific data such as SI-NET-seq and nascONT-seq. Due to the single-nucleotide resolution of SI-NET-seq data, we set the bin size to 1. We removed 0.1% of the strongest SI-NET-seq signal that we considered as outliers.

3′-RNA cleavage assay

The cleavage assay was designed similarly to an assay by Davidson et al. (2014). The main 3′ end cleavage sites of *MBNL1*, *PVT1* and *CDC42* were extracted from the database PolyA_DB v3.2 (Wang et al., 2018). Primer pairs that span the pA site as well as primer pairs upstream and downstream of the pA site were designed using Primer3 (Untergasser et al., 2012) and tested for specificity using NCBI Primer BLAST (Ye et al., 2012). A complete list of primer sequences is given in Table S7. Primer efficiency and specificity were tested using a template concentration series and by melting curve analysis, respectively.

After 2 h of treatment with dBET6, dTAG7 or DMSO, K562 cells were fractionated and subjected to chromatin RNA extraction as described in the 'SI-NET-seq' section. cDNA was generated from 500 ng chromatin RNA using the SuperScript III kit (Invitrogen) following the manufacturer's instructions with one modification. Random hexamer primers were used instead of oligo dT primers. 5 ng cDNA were used for each 10 μl RT-qPCR reaction. Treatment was performed for three (dBET6) or six (dTAG7) biological replicates. qPCR measurements were conducted in technical triplicates or quadruplicates. To minimize normalization artifacts, C_t values were normalized to the C_t values obtained for the DMSO control that were obtained for the same primer pairs, or to the C_t values of the most upstream gene-body position. Significance was determined by ANOVA using the GraphPad Prism software.

Proteomics analysis

Sample preparation for immunoprecipitation after formaldehyde crosslinking (IP-MS)

Sample preparation was done as described by Mohammed et al. (2016). Briefly, 5×10^7 K562 dTAG-BRD4 cells were crosslinked in FBS-free RPMI containing 1% methanol-free formaldehyde at room temperature for 8 min. Quenching was done by adding glycine to a final concentration of 250 mM and brief inversion. Cells were washed twice with ice-cold PBS and frozen in liquid nitrogen. Cell lysis was performed as described by Mohammed et al. (2016). To solubilize the chromatin, the samples were sonicated in TPX tubes (Diagenode) using five sonication cycles of 30 s on and 60 s off at high intensity, and at 4°C using a Bioruptor Plus sonicator (Diagenode). After adding 1/10 volumes Triton X-100, samples were cleared by centrifugation. For each sample, 300 μg magnetic Pierce protein A/G beads (Thermo Scientific) were blocked with 0.5% BSA/PBS and incubated for 1 h at room temperature with 6.7 μg antibody recognizing the HA epitope which is part of the dTAG. Cell lysates were incubated with the bead-coupled antibodies on a rotator overnight at 4°C. On the next day, beads were washed eight times with modified RIPA (50 mM HEPES, 0.5% (vol/vol) NP-40, 0.4% sodium deoxycholate, 1 mM EDTA, 300 mM LiCl, 1X protease inhibitor cocktail) and three times with 100 mM ammonium bicarbonate. Five or four biological replicate measurements were performed for K562 dTAG-BRD4 and untagged K562 cells, respectively.

Treatment and cell fractionation for quantitative chromatin mass spectrometry (chromatin-MS)

K562 cells were cultured in RPMI containing 10% dialyzed FBS (GIBCO) and 5% penicillin-streptomycin, to which either $^{13}C_6{}^{15}N_2$ L-lysine-2HCl and $^{13}C_6{}^{15}N_4$ L-arginine- HCl ('heavy RPMI'; Thermo Scientific), or L-lysine and L-arginine containing only light isotopes ('light RPMI'), were added. After at least five cell division cycles, the degree of the heavy isotope incorporation was checked as described by Ong and Mann (2007). For chromatin analysis, four replicates, including two label-swap replicates, were prepared. 5×10^6 SILAC-labeled K562 cells were treated with 500 nM dTAG7 or 100 nM dBET6 or with an equivalent volume of DMSO for 2 h and pooled after one PBS wash. Cell fractionation was performed as described for the SI-NET-seq approach using buffers containing phosphatase and protease inhibitors but no α-amanitin and RNase inhibitor. To solubilize the chromatin pellet, 750 U benzo-

Please cite this article in press as: Arnold et al., A BRD4-mediated elongation control point primes transcribing RNA polymerase II for 3′-processing and termination, Molecular Cell (2021), https://doi.org/10.1016/j.molcel.2021.06.026

nase (Sigma) were added, followed by 10 min incubation on ice and 5 min of agitation at room temperature. The solubilized chromatin was precipitated using the ProteoExtract protein precipitation kit (Merck Millipore) according to the manufacturer's instructions.

Treatment and cell lysis for quantitative total cell mass spectrometry

After 2 h treatment with dTAG7 or DMSO, 2 × 10^6 SILAC-labeled K562 dTAG-BRD4 cells were collected, washed with PBS and combined. A label-swap experiment was done in parallel. In total, four replicates were prepared. Each cell pellet was resuspended in 100 μL lysis buffer (8 M urea, 50 mM Tris pH 7.5, protease inhibitor cocktail, phosphatase inhibitor cocktail). The samples were vigorously shaken on a vortexing rotator for 30 min in the presence of 250 U benzonase. After complete solubilization, the samples were precipitated using the ProteoExtract kit (Merck Millipore) according to the manufacturer's instructions and snap-frozen as dry protein pellets.

Preparation for liquid chromatography mass spectrometry (LC-MS/MS)

For the IP samples, 30 μL of 100 mM ammonium bicarbonate was added to the magnetic beads followed by 'on-beads' digestion with 100 ng trypsin (Roche), shaking at 500 rpm at 37°C for 18 h. A second digest was performed using the same conditions for 4 h. The enzymatic reaction was stopped by adding formic acid to a final concentration of 5%. Peptide desalting using Pierce C18 tips (Thermo Scientific) was performed according to the manufacturer's instructions. Desalted peptides were reconstituted in 5% acetonitrile and 2% formic acid in H_2O, briefly vortexed, and sonicated in a water bath for 30 s prior to the injection for nano-LC-MS measurement.

The chromatin and the total cell samples were lysed under denaturing conditions in 300 μL of a buffer containing 3 M guanidinium chloride (GdmCl), 5 mM tris(2-carboxyethyl)phosphine, 20 mM chloroacetamide and 50 mM Tris-HCl pH 8.5. Lysates were denatured at 95°C for 10 min shaking at 1000 rpm in a thermal shaker and sonicated in a water bath for 10 min. A small aliquot of cell lysate was used for the bicinchoninic acid (BCA) assay to quantify the protein concentration. 30 μg protein of each lysate was diluted with the dilution buffer (10% acetonitrile, 25 mM Tris-HCl, pH 8.0), to a final concentration of 1 M GdmCl. Next, proteins were digested with LysC (Roche; enzyme to protein ratio 1:50, MS-grade) shaking at 700 rpm at 37°C for 2 h. The digestion mixture was diluted again with the same dilution buffer to reach 0.5 M GdmCl, followed by a tryptic digestion (Roche, enzyme to protein ratio 1:50, MS-grade) and incubation at 37°C overnight in a thermal shaker at 700 rpm. Peptide desalting was performed as above. Desalted peptides were reconstituted in 0.1% formic acid and 27 μg of the digest was further separated into four fractions by strong cation exchange chromatography (SCX, 3M Purification). Eluates were first dried in a SpeedVac, then dissolved in 5% acetonitrile and 2% formic acid in water, briefly vortexed, and sonicated in a water bath for 30 s prior to injection to nano-LC-MS.

LC-MS/MS instrument settings for shotgun proteome profiling and data analysis

LC–MS/MS was carried out by nanoflow reverse-phase liquid chromatography (Dionex Ultimate 3000, Thermo Scientific) coupled online to a Q-Exactive HF Orbitrap mass spectrometer (Thermo Scientific). Briefly, LC separation was performed using a PicoFrit analytical column (75 μm ID × 55 cm long, 15 μm Tip ID; New Objectives) packed in-house with 3-μm C18 resin (Reprosil-AQ Pur, Dr. Maisch). Peptides were eluted using a gradient from 3.8 to 40% solvent B in solvent A over 65 min at a 266 nL per minute flow rate for the IPs and a 2 h gradient per fraction was used for the chromatin samples. Solvent A was 0.1% formic acid, and solvent B was 79.9% acetonitrile, 20% H2O, and 0.1% formic acid. Nanoelectrospray was generated by applying 3.5 kV. A cycle of one full Fourier transformation scan mass spectrum (300−1750 m/z, resolution of 60,000 at m/z 200, AGC target 1E6) was followed by 12 data-dependent MS/MS scans (resolution of 30,000, AGC target 5E5) with a normalized collision energy of 25 eV. To avoid repeated sequencing of the same peptides, a dynamic exclusion window of 3 s for the IP and 30 s for the chromatin samples were used. Additionally, only peptides with charge states between two and eight were sequenced.

Proteomics data processing

Raw MS data were processed using the MaxQuant software (version 1.6.0.1) (Cox and Mann, 2008). All fractions of a sample were allocated to the corresponding replicate and analyzed jointly and searched against the human proteome database UniProtKB with 21,074 entries, released in 12/2018 for each project. The following parameter settings of MaxQuant were used for database searching: a false discovery rate (FDR) of 0.01 for proteins and peptides, a minimum peptide length of seven amino acids, and a mass tolerance of 4.5 ppm for precursor and 20 ppm for fragment ions. A maximum of two missed cleavages was allowed for the tryptic digest. Cysteine carbamidomethylation was set as a fixed modification, while N-terminal acetylation and methionine oxidation were set as variable modifications. In addition, the following SILAC labels were used for the chromatin and the total cell samples: $^{13}C_6{}^{15}N_4$-arginine and $^{13}C_6{}^{15}N_2$-lysine. The obtained output files of MaxQuant, which contained the peptide and protein identifications, accession numbers, percent sequence coverage of the protein, q-values and label-free quantification (LFQ) intensities or SILAC ratios, can be found in Table S3.

IP-mass spectrometry samples were analyzed using the Perseus software platform (version 1.6.2.3) (Tyanova et al., 2016). To increase sensitivity, the minimum number of common peptides used for pairwise protein ratio calculation was set to 1, and identifications were matched between runs. After filtering for potential contaminants, proteins matching the reverse sequence database or proteins for which only modified peptides were detected, LFQ intensities were log2 transformed and grouped based on the antibody used. Next, proteins that were not present in at least 70% of the replicates of the same sample type were removed. Missing values were imputed from a normal distribution of the measured values of each replicate, with the width set to 0.3 and a down-shift of 1.8. A two-sided two-sample Student's t test with Benjamini-Hochberg FDR = 0.05 was conducted for the HA IP and the HA control samples.

Please cite this article in press as: Arnold et al., A BRD4-mediated elongation control point primes transcribing RNA polymerase II for 3′-processing and termination, Molecular Cell (2021), https://doi.org/10.1016/j.molcel.2021.06.026

The analysis of the chromatin samples and of the total cell samples to determine specificity of the dTAG7 treatment was carried out based on normalized SILAC ratios that were calculated based on $\geq$ 2 common peptides using the Perseus software. To increase sensitivity, a supplemental analysis was performed based on $\geq$ 1 common peptide, as indicated. After filtering as described above, 1/x transformation of the ratios of the label-swap replicates and log2 transformation were computed. Proteins for which the corresponding ratios were missing in > 30% of the replicates were excluded from the analysis. Significant changes were identified by a two-sided one-sample Student's t test.

Statistical overrepresentation analysis of 'GO-slim Biological Process' terms was performed using PANTHER (version 15) (Mi et al., 2019). Significance was calculated by Fisher's exact test. Largely overlapping terms were excluded for visualization.

Sample preparation for native immunoprecipitation (IP)

4×10^7 K562 dTAG-BRD4 cells were collected and washed twice with ice-cold PBS. Cell pellets were incubated in 450 μl native IP buffer (20 mM Tris pH 8.0, 120 mM NaCl, 0.5% (vol/vol) NP-40, protease inhibitor cocktail, phosphatase inhibitor cocktail; antibodies against the HA-tag and SPT6 required 10% (vol/vol) glycerol, 50 mM NaCl) for 4 min on ice, followed by centrifugation for 10 min at 13,000 rpm and 4°C. Nuclei pellets were resuspended in 410 μl native IP buffer supplemented with 1000 U benzonase, incubated on ice for 40–120 min with repeated mixing and centrifuged for 10 min at 14,000 x g and 4°C. 10% of the sample volume was saved as input. 0.8 μM EDTA were added to the lysate. Next, the lysates were incubated with 2–10 μg antibody or isotype-matched pre-immune IgG for 4 h on a rotating wheel at 4°C, and for additional 1.5 h in the presence of 150 μg Protein G Dynabeads. Following the IP, beads were washed five times with 1 mL native IP buffer and proteins were eluted in 1X SDS buffer (50 mM Tris-HCl pH 7.0, 2% (wt/vol) SDS, 10% glycerol, 5% β-mercaptoethanol) at room temperature. Immunoblotting was performed as described above.

QUANTIFICATION AND STATISTICAL ANALYSIS

Statistical analysis was performed using built-in R and Perseus statistical tests. Statistical information including n, median and statistical significance values are indicated in the figure, text or the figure legends. Error bars in the experiments represent standard deviation of the mean values from either independent experiments or samples. Boxplot representations show the first quartile, median, third quartile and whiskers which extend to 1.5X of the interquartile range. Outliers are not shown. Details about experiment quantification can be found in STAR Methods. Unless indicated otherwise, statistical tests were performed two-sided and significance was defined as p value < 0.05.