SF3B1-targeted Splicing Inhibition Triggers Global Alterations in Transcriptional Dynamics and R-Loop Metabolism

Daisy Castillo-Guzman*, Stella R. Hartono*, Lionel A. Sanz, and Frédéric Chédin‡

Department of Molecular and Cellular Biology and Genome Center, University of California, Davis, Davis, CA 95616

* These authors contributed equally

‡ Corresponding author: flchedin@ucdavis.edu

SUMMARY

Efficient co-transcriptional splicing is thought to suppress the formation of genome-destabilizing R-loops upon interaction between nascent RNA and the DNA template. Inhibition of the SF3B splicing complex using Pladienolide B (PladB) in human K562 cells caused widespread intron retention and nearly 2,000 instances of R-loops gains. However, only minimal overlap existed between these events, arguing that unspliced introns do not cause excessive R-loops. R-loop gains were instead driven by readthrough transcription resulting from loss of transcription termination over a subset of stress-response genes, defining a new class of aberrant "downstream of genes" (DoG) R-loops. Such DoG R-loops were temporally and spatially uncoupled from loci experiencing DNA damage. Unexpectedly, the predominant response to splicing inhibition was a global R-loop loss resulting from accumulation of promoter-proximal paused RNA polymerases and defective elongation. Thus, SF3B1-targeted splicing inhibition triggered profound alterations in transcriptional dynamics, leading to unexpected disruptions in the global R-loop landscape.

HIGHLIGHTS

- Intron retention caused by SF3B1 inhibition does not lead to excessive R-loops
- A subset of genes shows readthrough transcription and accompanying R-loop gains
- SF3B1 inhibition causes broad reduction in nascent transcription and R-loop loss
- R-loop gains and DNA damage are temporally and spatially uncoupled

INTRODUCTION

During transcription, the nascent RNA can anneal to the DNA template strand behind the advancing RNA polymerase (RNAP), forming a stable RNA:DNA hybrid and causing the non-template DNA strand to loop out. The resulting non-B DNA structure, called an R-loop, is facilitated by favorable DNA sequence and negative superhelicity (Chedin and Benham, 2020; Drolet et al., 2003; Stolz et al., 2019). R-loops have been described from bacteria to plants to mammals (Santos-Pereira and Aguilera, 2015) and represent a prevalent class of alternative DNA structures (Chedin, 2016). Studies over the last decade have implicated R-loop formation in a variety of physiological processes such as transcription termination (Proudfoot, 2016; Sanz et al., 2016; Skourti-Stathaki et al., 2011), chromatin patterning and gene expression control (Chedin, 2016; Niehrs and Luke, 2020; Sanz et al., 2016; Skourti-Stathaki et al., 2019; Tan-Wong et al., 2019), and class switch recombination (Yu et al., 2003; Yu and Lieber, 2019). By contrast, so-called aberrant R-loops formed under a variety of pathological conditions have been extensively linked to phenomena of genome instability (Crossley et al., 2019; Garcia-Muse and Aguilera, 2019; Hamperl and Cimprich, 2014). Not surprisingly, a number of human diseases have recently been linked to altered R-loop metabolism (Richard and Manley, 2017). Despite the rising importance of R-loops in adaptive and maladaptive processes, many questions remain regarding the manner by which they are formed and regulated.

It is widely accepted that R-loops form co-transcriptionally upon re-invasion of the nascent RNA *in cis*. Interestingly, alterations of a number of co-transcriptional processes such as splicing, RNA export, cleavage and polyadenylation, have been linked to phenomena of R-loop-mediated genome instability from yeast to human cells (Aguilera, 2005; Aguilera and Garcia-Muse, 2012; Chan et al., 2014; Huertas and Aguilera, 2003; Li and Manley, 2005, 2006; Paulsen et al., 2009; Stirling et al., 2012). Earlier work suggested that inactivation in the splicing

regulator SRSF1 triggered the accumulation of R-loops, presumably upon formation of excessive R-loops via newly unspliced transcript portions (Tresini et al., 2015). SRSF1-depleted cells also showed accumulation of double-stranded DNA breaks (DSBs), a hypermutagenic phenotype, G2 arrest, and loss of cell viability (Li and Manley, 2005). Importantly, these phenotypes could be reversed upon over-expression of ribonuclease H1 (RNase H1), an enzyme with the clear biochemical ability to degrade RNA in the context of RNA:DNA hybrids (Cerritelli and Crouch, 2009). This suggested that excessive R-loop formation is directly involved in driving genome instability, and that splicing factors are responsible for limiting R-loop formation. Systematic genetic screens in human cells identified splicing-related genes as the top category of factors involved in the prevention of R-loop-mediated genome instability, as scored by the phosphorylated H2AX (γH2AX) DNA damage marker (Paulsen et al., 2009). Complementary work in yeast further underscored that introns attenuate R-loop formation and transcription-associated genetic instability via the recruitment of the spliceosome onto the pre-mRNA (Bonnet et al., 2017). Altogether, these studies suggest that proper splicing, export, and packaging of the nascent RNA into ribonucleoprotein particles contributes significantly to the maintenance of genome stability by keeping the RNA away from the DNA template and avoiding deleterious R-loop formation.

The nature and distribution of excessive R-loops expected to result from the disruption of co-transcriptional RNA processing have not been established at genome scale. Thus, the spatial relationship between putative aberrant R-loop gains and events of genome instability has never been directly tested. To address these gaps, we focused on splicing as a key co-transcriptional process linked to R-loop regulation. SF3B1 is a conserved and essential core subunit of the SF3B complex, a key component of the U2 spliceosome involved in branch site recognition and selection during pre-mRNA splicing (Sun, 2020). SF3B1 can be specifically inhibited via pharmacological compounds, including Pladienolide B (PladB) (Cretu et al., 2018;

Kotake et al., 2007; Yokoi et al., 2011). PladB and other macrolide-type splicing inhibitors rapidly abrogate splicing, causing a vast increase in events of intron retention (IR) (Boswell et al., 2017; Carvalho et al., 2017; Kashyap et al., 2015). PladB treatment provides a unique opportunity to determine whether unspliced introns can drive excessive R-loop formation and associated DNA damage, which is expected if the newly retained intron sequences form genome-destabilizing R-loops. This pharmacological approach further allowed us to temporally resolve the resulting changes in R-loop distribution, splicing events, and transcriptional dynamics, providing a unique view into the cellular response to acute splicing inhibition.

RESULTS

PladB treatment causes broad intron retention but little R-loop gains in human K562 cells.

We used RNA-seq to characterize alterations in splicing patterns resulting from PladB treatment two and four hours after drug application. Human K562 cells either mock-treated with DMSO or untreated were used as controls and two biological replicates were analyzed at each time point (Figure 1A). As expected, PladB treatment caused broad, time-dependent, splicing alterations dominated by events of intron retention (IR) and to a lesser extent, skipped exons (SE) (Figure 1B). Validation of RNA-seq analysis by RT-qPCR at the known PladB target gene *DNAJB1* (Kashyap et al., 2015) confirmed that IR gradually accumulated over time and could be detected as soon as 15 minutes after treatment (Figure S1A,B). IR was observed at *DNAJB1* in strand-specific RNA-seq analysis of poly(A)-tailed mRNA pools two- and four-hours post treatment (Figure 1C). In total, over 7,000 independent IR events were identified in human K562 cells four hours post PladB treatment. Such retained introns provide an ideal cohort to test the hypothesis that splicing inhibition causes gains of R-loops over regions now associated with unspliced introns. To determine how R-loops respond to PladB-mediated splicing inhibition, we profiled R-loop distribution using DRIP-seq (Sanz and Chedin, 2019) at the same time points. Only a modest number of R-loop gains (RLGs) were observed post PladB treatment with 1,538 and 1,877 peaks identified within the limits of our statistical thresholds two and four hours post-PladB, respectively (Figure 1D). Unexpectedly, the vast majority of significant R-loop changes corresponded to events of R-loop loss, highlighting 18,518 and 28,836 peaks two and four hours post-PladB, respectively. To determine whether IR associates with RLGs, we first focused on individual loci that showed clear retention of introns. At *DNAJB1*, no significant RLGs could be observed over the retained intron (Figure 1E). Similar observations were made over a retained intron at the *RPL13A* gene (Figure S1C). More broadly, less than 2% of loci with intron

retention events intersected with RLG peaks (121 **/** 7289; Figure 1F). This suggests that contrary to expectations, large-scale intron retention caused by acute PladB-mediated splicing inhibition, did not cause R-loop gains over unspliced transcripts.

Splicing inhibition is associated with readthrough transcription and accompanying R-loops at a subset of genes.

To understand the nature of the RLGs caused by PladB treatment, we first determined where these peaks mapped in the genome. 77% of R-loop peaks in untreated K562 cells mapped to genic regions, as expected (Sanz et al., 2016). By contrast, nearly two thirds of the RLGs in PladB-treated cells mapped to intergenic regions (Figure 2A) indicating that PladB treatment unexpectedly led to RLGs over non-genic regions. To assess how such RLGs arose, we focused on a region with numerous clustered peaks of RLGs located downstream of the *VAPA* gene (Figure 2B). The peaks extended from the 3'-end of *VAPA* and grew directionally with transcription in a time-dependent manner. After two hours, the edge of the wave of RLGs extended 135 kb downstream of the *VAPA* gene. After four hours, a trail of R-loops was detected 210 kb downstream of *VAPA*. DRIP-qPCR on three additional replicates confirmed these findings and showed that R-loops increased 20-fold 60 kb downstream of the *VAPA* poly(A) site (PAS) two hours post PladB treatment and 6-fold 200 kb downstream of *VAPA* four hours post treatment (Figure 2C). This pattern is consistent with a directional and time-dependent propagation of a wave of R-loops from the *VAPA* gene. Analysis of R-loops using a high-resolution, strand-specific, iteration of DRIP-seq (sDRIP-seq; see methods) confirmed that the RLGs were stranded and co-directional with *VAPA* transcription (Figure S2A). Similar findings were observed for two additional genes, *RPL9* and *CYCS* that showed a stranded increase in R-loop signal extending downstream of their PAS (Figure S2B-E).

Given that R-loops form co-transcriptionally, these observations suggest that PladB-mediated splicing inhibition caused readthrough transcription downstream of *VAPA*. RT-qPCR analysis of total RNA confirmed the existence of transcripts downstream of *VAPA* upon PladB treatment (Figure 2D). Transcripts were detected 15 and 60 kb downstream of *VAPA* 30 min post PladB treatment and steadily built over time up to a 35 to 60-fold increase over controls. 120 kb downstream of *VAPA*, transcripts were only detected 120 and 240 min after treatment, consistent with the steady progression of the RNA polymerase (RNAP) machinery from the 3'-end of *VAPA* post PladB treatment. To further explore the possibility that PladB treatment triggers downstream of genes (DoG) transcription, we analyzed nascent transcription in control and PladB-treated K562 cells using EU labeling and total RNA-seq (see Methods). Nascent transcripts were clearly observed downstream of *VAPA* and progressed unidirectionally as a function of time from the normal *VAPA* termination region (Figure 2E). Similar findings were observed at the *CYCS* and *RPL9* genes (Figure S2F,G). Thus PladB treatment appears to trigger readthrough transcription and accompanying co-transcriptional R-loops.

DoG transcription affects a specific subset of stress-responsive genes and is associated with broad *de novo* R-loop gains.

To determine how broad the PladB-induced transcriptional readthrough was, we systematically annotated whether events of RLGs correspond to DoG R-loops initiating from a neighboring gene (Methods). A total of 429 genes were identified with these characteristics, corresponding to the vast majority of RLGs (1,505/1,877 peaks; Figure S3A). Metaplots confirmed increased R-loops and nascent transcription extending downstream of the host genes PASs for a distance of 50-75 kb (Figure 3A, B). Annotation of individual DoG R-loop lengths confirmed that the median collective length of DoG R-loops was 49.2 and 50 kb two and four hours post PladB treatment, respectively (Figure 3C). This was significantly higher than the length of terminal R-loops recorded for the same genes under control conditions. Interestingly,

PladB-sensitive DoG genes showed longer terminal R-loops than expression-matched genes even under control conditions (18.2 kb versus 31.5 kb) indicating that DoG genes may have an intrinsic propensity to terminate further downstream of their PASs (Figure 3C). Overall, PladB-mediated splicing inhibition was associated with DoG transcription and accompanying DoG R-loops for a large subset of genes. Collectively, DoG R-loops observed four hours after PladB treatment covered 2.81 megabases of genomic space, defining the first example of a class of excessive R-loops identified at genome-scale. DoG genes showed significant enrichments for specific gene ontologies (GOs), including translation initiation and ribosome biogenesis, viral process, response to stress, mitochondrial gene expression, and chromatin organization (Figure 3D, Figure S3B). The appearance of stress-related and viral-related GOs is notable given precedents for DoG transcription triggered upon environmental stresses and viral infections (Bauer et al., 2018; Cardiello et al., 2018; Hennig et al., 2018; Vilborg et al., 2015; Vilborg et al., 2017).

Global R-loop losses are caused by PladB-induced negative feedback on transcription elongation.

Unexpectedly, the overwhelming majority of R-loop changes in response to PladB (28,836 / 30,713 events or 93.8% 4 hours post PladB) corresponded to R-loop losses (Figure 1D). To understand the mechanisms driving these events, we first asked where R-loop losses (RLLs) occurred. Contrary to RLGs, RLLs matched primarily to transcribed gene bodies (44%) and terminal genic regions (29%) (Figure 4A). An additional 20% of RLLs matched to normally transcribed regions located immediately downstream of terminal PAS sites. By contrast, only a small minority of RLLs mapped to promoter and promoter-proximal regions. To further understand how these losses originated, we focused on the representative R-loop-prone *DGCR2* gene (Figure 4B). RLLs were clearly visible in the gene body and gene terminus, with little initial impact around the promoter region and early first intron. These RLLs were validated

by DRIP-qPCR in three additional independent replicates, which confirmed that early R-loops were unchanged but R-loop formation along the gene body was progressively reduced from 10-fold in the middle of the gene to 50-fold by the gene terminus (Figure 4C). The simplest hypothesis to account for these observations is that transcription elongation was impaired, leading to a directional reduction of co-transcriptional R-loops. To test this, we took advantage of our matched EU-seq datasets and observed that nascent transcription along *DGCR2* was significantly affected, with EU incorporation only located in the early portion of intron 1, where R-loops still occurred (Figure 4D). Similar observations were made for other genes (Figure S4A-C). We note that for long genes (ex: *CHCHD3* (293 kb), Figure S4A), RLLs at the terminus of the gene were only fully observed after four hours of PladB treatment (Figure S4D), suggesting that RLLs arise from a progressive loss of elongation capacity from newly engaged RNA polymerase complexes.

We systematically annotated genes showing significant RLL in their gene body or terminal regions and identified 5,651 genes. As expected from single gene examples, metaplot analyses confirmed loss of R-loops gradually occurred along gene bodies and was maximal around gene termini (Figure 4E). Clear loss of nascent transcription was also observed along these genes (Figure 4F). In general, nascent transcription and R-loops were in strong agreement: 77% of genes annotated as losing nascent transcription also showed RLL (Figure S4E). The remainder of genes showing EU-seq loss without accompanying RLL simply had little to no R-loop loads under control conditions. Conversely, only about half the RLL genes detected using DRIP-seq signal also showed loss of EU-seq signal, suggesting that co-transcriptional R-loop formation is highly sensitive to alterations in nascent transcription above what can be measured from the relatively low EU-seq signal obtained after a brief 20 min EU pulse. Overall, this data suggests that splicing inhibition triggered profound negative feedback on transcription elongation.

PladB treatment leads to increased promoter-proximal pausing and defective elongation.

A large subset of RLL genes, termed thereafter EL genes, showed early elongation loss and R-loop loss. A smaller group of genes were instead characterized by DoG transcription and associated R-loop gains. A third class of transcribed genes (n=5,455) showed neither loss of transcription or R-loops in their gene bodies, nor gains of DoG R-loops, and were arbitrarily defined as “unaffected”. Thus, PladB treatment resulted in broad and contrasting alterations in transcriptional dynamics. We next asked if these gene classes possessed specific characteristics that might illuminate their responses to PladB.

Length-wise, unaffected genes were significantly shorter than EL and DoG genes while DoG genes were located significantly further away from their nearest upstream and downstream neighbors compared to other gene classes (Figure S5A,B), suggesting they are located in less gene-dense neighborhoods. DoG genes showed strikingly higher nascent transcription around their promoter regions compared to both unaffected and EL genes in the absence of PladB (Figure 5A). A pronounced peak was observed immediately downstream of the transcription start site (TSS), reflecting promoter-proximal pausing, and extended in a broader band corresponding to the first 500-1,000 bp of promoter-downstream sequences. This suggests that DoG genes are heavily regulated by pause-release mechanisms and carry high loads of paused RNAPs. Importantly, PladB treatment caused a further significant increase in promoter-proximal pausing for both DoG and unaffected genes, especially two hours post PladB treatment. EL genes, which had the lowest pausing under control conditions, showed the least gains upon PladB treatment (Figure 5A). Thus, the three gene classes identified here showed important differences in their amounts of promoter-proximal paused RNAP complexes. Traveling ratios (TRs), which are often used to identify paused genes (Rahl et al., 2010) confirmed that DoG genes showed significantly higher TRs compared to both unaffected and EL genes under control conditions (Figure S5C,D). Interestingly, the relative increase in active elongating RNAP

measured 2 kb downstream of the TSS was significantly lower than the TSS-proximal increase of paused polymerases (Figure 5A). This indicates that a high proportion of RNAP complexes newly recruited to promoters upon PladB treatment were not released into elongation.

We envisioned that the increased RNAP recruitment to promoters would lead to increased levels of nascent transcription throughout gene bodies. Analysis of EU-seq patterns, instead, suggest that elongation in the presence of PladB was defective. To visualize this, we plotted the gene body ratios of EU-seq signals observed two and four hours after PladB treatment over the EU-seq signal for the same genes under control conditions (Figure 5B). All three gene classes showed an increased EU-seq ratio upon PladB treatment at the beginning of gene bodies, consistent with the increased RNAP recruitment at the promoter and the progressive release of these polymerases into elongation. However, the ratios showed a clear negative slope, indicating that nascent transcription levels became gradually reduced through elongation. For all three gene classes, the ratios ultimately fell below that of control cells, but at noticeably distinct points. For DoG genes, which had the highest promoter-proximal RNAP loads, the ratios didn't fall below control until about 85% through gene bodies. Unaffected genes showed the same negative slope and crossed below control levels around 70% of their gene bodies. EL genes, which had the lowest RNAP promoter loads under control conditions and the smallest increase upon PladB treatment, fell below control levels early (5-12%) within gene bodies. This indicates that while transcription elongation was broadly affected for all gene classes, EL genes were disproportionally sensitive owing to the lack of a robust store of promoter-proximal paused RNA polymerases.

Defective elongation is not due to global premature termination.

Inhibition of U1 small nuclear ribonucleoprotein (snRNP) triggers a strong negative feedback on transcription caused by premature termination via the cleavage and

polyadenylation complex (Kaida et al., 2010). We investigated if the loss of transcription elongation observed here shared a similar mechanism by looking for telltale signs of premature termination, namely the formation of short polyadenylated transcripts resembling early 3'UTRs in RNA-seq data. While a few instances of such alternative polyadenylation sites were found (Figure S5E), they remained rare. Systematic annotation of RNA-seq data for gains of intronic sequence reads revealed that out of the 5,651 EL genes, 2,421 (42.8%) carried significant intron read gains. In 85% of cases, such intronic read gains corresponded to intron retention events, with a small minority annotated as possible alternative 3'UTRs (Figure 5C). Inspection of EU-seq reads for the presence of polyadenylated transcripts failed to reveal extensive polyadenylation of nascent transcripts. Upon U1 snRNP inhibition, transcription termination occurs typically in the first intron (Berg et al., 2012; Kaida et al., 2010; Oh et al., 2017). Using DRIP-seq which provides a sensitive readout of nascent transcription, we annotated the position of the first detectable RLL in the gene bodies of EL genes. While elongation loss was preferentially observed early on in gene bodies (Figure S5F), losses were not confined to early introns and many individual events were instead spread out. We therefore conclude that unlike U1 snRNP, the loss of elongation capacity induced by PladB is unlikely to involve premature termination.

DoG genes preferentially show distal splicing alterations which may be coupled to readthrough transcription.

At first glance, DoG and unaffected genes responded similarly to PladB; yet clear readthrough transcription and DoG R-loops were only seen for DoG genes. Since proper transcription termination has been linked to splicing, especially terminal exon splicing (Davidson and West, 2013; Herzel et al., 2017; Rigo and Martinson, 2008, 2009), we analyzed whether DoG genes showed a distinct pattern of splicing alterations compared to unaffected and EL genes. We first focused on IR since it defined the largest class of aberrant splicing events. EL

genes showed predominantly IR events early in their gene body, consistent with elongation being restricted (Figure 5D). Unaffected genes also underwent IR events predominantly early. In contrast, DoG genes were significantly less likely to exhibit IR early and instead IR was spread throughout, often affecting terminal or near terminal introns. A similar trend with an even sharper contrast between DoG and unaffected genes was obtained when analyzing skipped exon events (Figure S5G). We note that no single exon gene was present in the DoG gene dataset even though 1,247 single exon genes were expressed in K562 cells and 40 were expected to display DoG transcription at random. These results suggest that DoG genes are significantly more predisposed to terminal aberrant splicing events.

We also wondered if DoG genes might display characteristics predisposing them to weaker transcription termination efficiencies. Analysis of PAS motifs failed to reveal significant differences between gene classes. However, we noted that DoG genes showed unusual R-loop patterns through their gene bodies under control conditions. As expected from their high loads of promoter-proximal RNAPs, DoG genes showed significantly higher R-loop levels over promoter regions and early gene bodies (Figure 5E). However, DoG genes showed significantly reduced R-loop levels towards gene ends compared to both unaffected and EL genes. Unaffected genes, in particular, showed a clear rise in terminal R-loops. Importantly, terminal R-loops have been associated with efficient transcription termination (Proudfoot, 2016; Sanz et al., 2016), indicating that DoG genes may possess intrinsically poorer termination mechanisms and therefore may be more prone to DoG transcription. Indeed, terminal R-loops were longer for DoG genes than unaffected genes even under control conditions (Figure 3C).

PladB induces a strong cellular stress response accompanied by reduced expression of DNA damage response genes.

We used RNA-seq to measure changes to mRNA pools following PladB treatment. 721 genes showed significant increases in gene expression (Figure 6A). Importantly, these up-regulated genes were dramatically enriched for a few, highly congruent, biological functions, including protein translation and ribosome biogenesis, RNA metabolic processes, and mitochondrial ATP synthesis (Figure 6B, S6A). Similar ontologies were observed for DoG genes (Figure 3E), suggesting that R-loop gains reflected at least in part an increase in expression. The strong response towards increased protein translation capacity is consistent with a cellular adaptation to dwindling mRNA pools due to transcription elongation loss. Increased energy production via mitochondrial respiration may represent a second cellular compensatory mechanism.

1,180 genes were significantly down-regulated (Figure 6A), far less than the number of genes that showed loss of nascent transcription. This suggests that the majority of mRNA pools remained stable at least four hours post splicing inhibition. Nonetheless, 62% of down-regulated genes also showed loss of nascent transcription by DRIP-seq or EU-seq (Figure S6B), suggesting that loss of transcription capacity was starting to affect a subset of mRNAs. GO analysis revealed that down-regulated genes were strongly enriched for genes involved in gene expression regulation and to a lower extent, for genes involved in cell cycle control, regulation of viral release, and the cellular response to DNA damage (Figure 6C, S6C). Among the top down-regulated genes in terms of gene expression and nascent transcription was *RECQL5* (Figure 6D, S6D), which encodes a transcription-associated elongation factor. Loss of RECQL5 leads to transcription stress, ultimately resulting in genome instability (Saponaro et al., 2014). More broadly, 37 genes implicated in the DNA damage response (DDR) were significantly downregulated 4 hours post PladB treatment (Table S1), including genes important for

homologous recombination pathway (XRCC2, XRCC3, FANCE), replication stress response (ATRIP), and nucleotide excision repair (XPC), among others. The number of DDR genes undergoing down-regulation increased to 300 and 242 genes when taking into account measures of nascent transcription using DRIP-seq and EU-seq, respectively (Table S1). Thus, PladB treatment leads to the down regulation of an important subset of genes, including a variety of DDR genes. Overall, PladB triggered the induction of a strong cellular stress response against the backdrop of a widespread loss of gene expression affecting numerous biological pathways, including the DNA damage response.

PladB treatment triggers mild DNA damage induction that is temporally and spatially uncoupled from DoG R-loop accumulation

PladB treatment was previously shown to trigger a mild DNA damage response and it has been suggested that this response was caused by elevated R-loop levels (Nguyen et al., 2019). We showed here that elevated R-loop levels are clearly visible over DoG genes two hours post-PladB treatment and that DoG transcripts can be detected as early as 30 minutes (Figures 2, 3). We therefore determined if DNA damage induction, as measured by γH2AX immunofluorescence staining, was temporally coupled to DoG R-loop formation. No significant γH2AX foci increase was observed four hours after PladB despite abundant evidence for DoG R-loops. A significant increase in γH2AX foci only occurred 24 hours post-PladB treatment (Figure 7 A,B), although it remained mild compared to the response to Etoposide (ETO) used here as a positive control. This suggests that the DNA damage response triggered by PladB is temporally uncoupled to DoG R-loop accumulation. We next asked if the loci experiencing increase γH2AX deposition would spatially overlap with DoG R-loops, as expected if such R-loops initiate DNA damage. For this, we performed chromatin immunoprecipitation (ChIP) experiments against γH2AX over multiple DoG R-loop hotspots. The DIvA/4-hydroxy tamoxifen

(4OHT) system was used to induce sequence-specific double strand breaks (DSBs) at predictable genomic loci (Iacovoni et al., 2010) as a positive control. ChIP-qPCR showed clear enrichment for γH2AX in 4OHT-treated cells compared untreated cells and to an IgG negative control (Figure 7C). By contrast, γH2AX was not increased over any locus four hours after PladB treatment. After 24 hours, while a trend towards increased γH2AX was observed, this trend was often not statistically significant and, crucially, was not unique to DoG R-loop loci. Indeed, multiple regions overlapping EL genes also showed a trend towards increased γH2AX (Figure 6C). Taken together, these results suggest that, although PladB treatment induces DNA damage, such damage does not spatially or temporally correlate with regions that accumulate excessive R-loops.

DISCUSSION

Intron retention does not drive R-loop formation

The retention of introns in pre-mRNAs upon splicing inhibition was hypothesized to lead to formation of excessive R-loops. Our results demonstrate that this hypothesis is incorrect, at least under the conditions tested here. PladB treatment caused the retention of over 7,000 introns in mRNA pools (Figure 1), a likely underestimate of the impact of SF3B1 inhibition on splicing at the nascent RNA level (Drexler et al., 2020; Nojima et al., 2015). Despite this, relatively few events of R-loop gains were identified genome-wide and less than 2 percent of annotated retained introns overlapped with R-loop gains. We therefore conclude that inhibiting the activity of the U2 spliceosome via targeting SF3B1 does not trigger R-loop gains over retained introns.

To explain these findings, we propose a model in which R-loop initiation can only occur over a short window located immediately upstream of the advancing RNA polymerase. The nascent RNA, once it exits the RNA polymerase, is proposed to sample the energy landscape (RNA-DNA base-pairing and DNA topology) such that initiation will occur only if the formation of a short nascent R-loop can withstand the energetic cost of forming a junction between B DNA and a three-stranded R-loop (Chedin and Benham, 2020). Once the RNA has passed this initiation window, it can no longer interact with the DNA template, whether or not it is spliced. We note that in cells, the nascent RNA will be rapidly bound by a host of RNA binding proteins (Metkar et al., 2018) and form additional secondary structures, further precluding its interaction with the template. This co-transcriptional R-loop formation model is consistent with most known aspects of R-loop distributions, including the observation that R-loops most often form over

introns and do not show any evidence that they are constrained by splicing junctions when profiled at the single-molecule level (Malig et al., 2020).

SF3B1-mediated splicing inhibition triggers global promoter-proximal RNAP pausing and elongation loss.

This study revealed that PladB treatment caused dramatic alterations in transcriptional dynamics, in addition to the expected splicing changes. The strongest response was a widespread loss of nascent transcription, detected initially as prominent R-loop losses over more than 5,000 genes (Figures 1, 4, 5). This loss was not accompanied by premature transcription termination associated with canonical cleavage and polyadenylation, as observed upon inhibition of the U1 small nuclear ribonucleoprotein (snRNP) (Berg et al., 2012; Kaida et al., 2010; Oh et al., 2017). Thus, U1 and U2 snRNP inhibition both caused pronounced transcription elongation loss, confirming the tight integration of gene expression programs with the splicing machinery through diverse feedback mechanisms (Bentley, 2014; Herzel et al., 2017; Hsin and Manley, 2012).

We show here that PladB treatment caused a sharp increase in promoter-proximal RNAP pausing. We further showed that a significant fraction of these paused but active complexes are not released into elongation, leading to a broad accumulation of nascent transcripts early in gene bodies for all gene classes (Figure 5). A similar accumulation of promoter-downstream non-polyadenylated transcripts was detected upon treatment of human HeLa cells with Isoginkgetin (IsoG) (Blasco, 2007), a non-SF3B splicing inhibitor that targets the U2 spliceosome at a later stage (O'Brien et al., 2008). These findings are consistent with the

notion that the release into elongation from a paused state is often the rate-limiting step for gene expression (Adelman and Lis, 2012; Core and Adelman, 2019; Mayer et al., 2017). Similarly, our work is in agreement with observations that prolonged pausing can lead to a reduction in release frequency and an overall decrease in RNA synthesis rates (Gressel et al., 2019; Gressel et al., 2017). Given these precedents, we suggest that SF3B1-targeted inhibition by PladB causes a global promoter-proximal pausing and a negative feedback on RNAP release from the paused state, leading to reduced elongation capacity and mRNA synthesis. The observation that PladB and other SF3B-targeted macrolides caused significant reduction in the levels of elongation-associated Serine 2 phosphorylated RNAP and the dissociation of RNAP from chromatin (Koga et al., 2015) is consistent with our observations. Additional work will be required to dissect the mechanistic basis by which PladB and other U2 snRNP-targeted splicing inhibitors negatively impact elongation release, as earlier studies suggest that it is distinct from a simple inhibition of the Serine 2-targeted CDK9 kinase (Boswell et al., 2017; Koga et al., 2015). In addition to causing enhanced RNAP pausing, PladB causes a further elongation defect in those complexes that have been released from the promoter. This was detected as a clear and progressive decrease in nascent transcription yields through gene bodies under PladB conditions (Figure 5B). We envision that this impaired elongation capacity could indirectly result from secondary perturbations in gene expression patterns, particularly given that the *RECQL5* elongation factor is one of the top down-regulated gene (Figure 6). Depletion of RECQL5 levels is expected to increase stalling and backtracking events (Saponaro et al., 2014), which could further compound the impact of enhanced promoter-proximal pausing on gene expression.

Interestingly, while the effect of PladB on transcriptional dynamics appears to apply uniformly to all genes, the resulting impact on gene expression varied depending on the intrinsic characteristics of the broad gene classes identified here (Figure 7D). DoG genes, which

possess large loads of paused RNAP early in their gene bodies even under control conditions, maintained sufficient elongation capacity for transcription complexes to complete their transcription. DoG genes were indeed broadly represented in a class of up-regulated stress-response transcripts (Figures 3, 6). Unaffected genes, by virtue of their short size and strong gains of RNAP upon PladB treatment, largely escaped elongation failure in response to PladB. EL genes, by contrast, possessed much reduced promoter-proximal active RNAP loads under control conditions and showed the lowest surge in RNAP pausing in response to PladB treatment. Due to the combination of these factors, EL genes suffered dramatic early elongation losses which dominated the PladB gene expression response.

PladB treatment causes readthrough transcription accompanied by *de novo* R-loop formation.

One additional effect of PladB treatment was the induction of DoG readthrough transcription for a subset of 429 genes (Figures 2, 3). We suggest that readthrough transcription may directly result from splicing alterations, consistent with the well-known links between splicing, particularly terminal exon splicing, and proper termination (Davidson and West, 2013; Kyburz et al., 2006; Rigo and Martinson, 2008, 2009). Readthrough transcription was also reported for HeLa cells treated with IsoG (Boswell et al., 2017), suggesting that reduced termination is a hallmark of U2 snRNP inhibition. Our study, however, clarified which gene class was vulnerable to such effects. DoG genes, in addition to escaping the pausing and elongation defects by virtue of their large stores of promoter RNAPs, also showed intrinsically lower termination efficiencies even under control conditions (Figure 3). This may be due to two intrinsic characteristics. First, DoG genes form only low levels of termination-associated R-loops compared to unaffected genes which are also short and transcribed (Figure 5). Second, DoG

genes show a propensity to undergo distal splicing alterations compared to unaffected genes, which may disproportionally impact the mechanistic coupling between U2 snRNP and termination complexes for this gene class.

Importantly, almost all events of R-loop gains observed upon PladB treatment mapped to DoG regions, delineating a novel class of aberrant *de novo* R-loops. Collectively, these excessive R-loops occupied nearly 3 megabases of genomic space. We note, however, that R-loop losses overwhelmingly dominated over R-loop gains, reflecting broad trends in nascent transcription. These findings contrast with previous studies which reported that PladB and other splicing inhibitors caused broad R-loops gains (Nguyen et al., 2018; Nguyen et al., 2017; Wan et al., 2015). Such claims, however, relied on S9.6-based imaging approaches which are unreliable without proper controls due to the affinity of the S9.6 antibody for structured RNAs (Hartono et al., 2018; Phillips et al., 2013; Smolka et al., 2020). The observations that PladB treatment triggers a broad enlargement of RNA-rich nuclear speckles (Effenberger et al., 2014) suggests that RNA distribution is profoundly altered, further raising the possibility of significant imaging artefacts.

PladB-induced transcriptional alterations resemble other stress responses.

DoG genes were enriched for GOs related to protein translation and response to cellular stress, consistent with our observation of a broad cellular response characterized by increased ribosome biogenesis, protein translation, mitochondrial respiration, and ATP synthesis (Figure 6). Interestingly, DoG transcription has been observed in response to a range of cellular stresses including oxidative stress (Giannakakis et al., 2015; Vilborg et al., 2017), heat shock

(Cardiello et al., 2018; Vilborg et al., 2017), and osmotic stress (Vilborg et al., 2015; Vilborg et al., 2017), in addition to viral infections (Bauer et al., 2018; Heinz et al., 2018; Hennig et al., 2018). However, the mechanisms underlying DoG transcription are not clear. Our observations raise the possibility that splicing disruptions leading to readthrough transcription represent a common molecular link underlying DoG transcription. In support of this, SF3B1 is a stress-sensitive splicing factor that responds to heat shock (Kim Guisbert and Guisbert, 2017). In addition, we note that these cellular stresses and viral infections also led to a broad repression of transcription and gene expression, as observed here upon splicing disruption. Heat shock, for instance, caused decreased RNAPII occupancy across most protein-coding genes (Cardiello et al., 2018), accumulation of promoter-proximal paused RNAP, and broad gene down-regulation (Gressel et al., 2019; Mahat et al., 2016). Our findings therefore suggest that splicing complexes may play an important role in sensing and responding to a variety of environmental and infectious stresses through the modulation of gene expression programs.

Rethinking the links between splicing inhibition, genome instability, and aberrant R-loops.

PladB treatment triggers a genome instability response that can be measured by the induction of γH2AX foci (Nguyen et al., 2017; Wan et al., 2015). The identification of a novel class of *de novo* DoG R-loops enabled us to test if these “excessive” R-loops are indeed driving genome instability, as commonly assumed. However, under our conditions, the DNA damage response was delayed and only detectable after 24 hours, when DoG transcription was observed as rapidly as 30 min (Figure 2). Considering that toxic R-loop intermediates might require time to be processed into DNA breaks, we next tested if the γH2AX DNA damage marker, would show enrichment over regions that accumulate excessive R-loops. However, this

was not the case. Instead, modest γH2AX gains were observed seemingly randomly at some, but not all, DoG sites tested and over EL genes showing clear R-loop losses. The DNA damage response observed upon PladB treatment was therefore temporally and spatially uncoupled from the accumulation of the common γH2AX marker, raising questions as to whether excessive R-loops represent toxic genomic intermediates. An alternative model for our observations is that the transcriptional losses driven by PladB led to the progressive down-regulation of DDR genes (Table S1), sensitizing cells to spontaneous genomic insults and ultimately resulting in DNA breaks. The strong down-regulation of *RECQL5* (Figure 6) may further enhance this phenomenon by causing transcription-associated DNA rearrangements (Saponaro et al., 2014). Such an indirect mechanism predicts that the resulting genome instability response should be temporally and spatially uncoupled from R-loop events, as observed here. We cannot rule out, however, that the increased R-loop loads driven here upon SF3B1 inhibition were not sufficient to mediate an effective R-loop-mediated DNA damage response. Analysis of additional cellular models of R-loop dysfunction will address this issue.

Previous studies have suggested that proper splicing is necessary to protect the genome from the damaging consequences of excessive R-loops (Li and Manley, 2005; Paulsen et al., 2009; Tresini et al., 2015). We show here that the predominant response of the R-loop landscape to an acute splicing inhibition is a global loss of R-loops through the gene bodies of thousands of genes. It is likely that this response is not unique to the SF3B1-targeted inhibition. Inhibition of the U1 spliceosome led to premature transcription termination globally (Kaida et al., 2010). U2, U4, and U6 snRNA depletion led to significant reduction in the levels of the elongating form of RNAPII (Koga et al., 2015), arguing that splicing inhibition at various stages of spliceosome assembly is likely to cause a similar global loss of gene expression accompanied by R-loop losses. The model that broad gains in “aberrant” genic R-loops drive

genome instability upon depletion of splicing factors might therefore need to be reconsidered. Our findings highlight the necessity of directly characterizing R-loop patterns at the genome scale and of taking into account the dynamic transcriptional responses that likely accompany any genetic or pharmacological perturbation, particularly when analyzing RNA processing and other nuclear factors.

FIGURE 1: A. Schematic of experimental setup, see text for details. **B.** Summary of PladB-mediated splicing alterations identified by RNA-seq. The number of skipped exons (SE), alternative 5' splice sites (A5SS), alternative 3' splice sites (A3SS) and intron retention (IR) gained or lost after 2 and 4 hours of PladB treatment are graphed. **C.** Genome browser screenshot over the representative *DNAJB1* gene showing plus and minus strand poly(A) RNA-seq signal obtained from controls and PladB treated K562 cells. Intron retention is observed two- and four-hours post PladB treatment (asterisk). **D.** XY plot of normalized DRIP-seq counts for control and treated (PladB 4 hours) samples. Each dot on the graph corresponds to a DRIP-seq peak. Red/blue colors indicate significant gains/losses of signals (adjusted p-value of < 0.05 and $|\log_2(FC)| \geq 1$). **E.** Genome browser screenshot showing the same region as in (C) but now displaying DRIP-seq signals. No increase in R-loop signal is detected over the retained intron. **F.** Venn diagrams displaying the intersection of events of intron retention and R-loop gains.

FIGURE 2: A. Location analysis of DRIP peak gains for control (left) and PladB-treated samples (right) across genomic compartments (color-coded according to the schematic below). R-loop distribution across compartments is indicated by percentages. TSS, transcription start site; PAS, polyadenylation site. **B.** Genome browser screenshot over the representative *VAPA* gene and ~200 kb downstream showing DRIP-seq signal from controls and PladB-treated K562 samples. R-loop gains occur directly downstream of gene (DoG - colored bar). **C.** Bar chart of DRIP-qPCR (as fold change over control) for *VAPA* and two regions downstream of PAS. Each bar is the average of three-independent experiments (shown with SE). **D.** Bar chart of RT-qPCR (as fold change over control) at indicated loci through a time course after PladB treatment. Each bar represents the average of three independent experiments (shown with SD). **E.** Genome browser screenshot over the same region as in (B) but now displaying EU-seq signals. Regions with increased R-loops also have increased nascent transcription.

FIGURE 3: A,B. Metaplots of DRIP-seq (top) and EU-seq (bottom) signals extending from nearest PAS for all annotated DoG genes. For each time point, the signal is shown as a trimmed mean (line) surrounded by SE (shaded). **C.** Terminal R-loop lengths (measured from DRIP-seq data) for DoG genes and expression-matched unaffected genes under control conditions and two- and four-hours post PladB. P-values were determined by a Wilcoxon Mann-Whitney test. The numbers below indicate median terminal R-loop lengths at the respective time points. **D.** Enriched gene ontologies for DoG genes. The dashed line indicates 1 % FDR.

FIGURE 4: A. Location analysis of R-loop losses for PladB-treated (4 hours) samples across genomic compartments. Numbers indicate the percentage occupied by each compartment. **B.** Genome browser screenshot over the representative *DGCR2* gene showing DRIP-seq signal from controls and PladB-treated K562 samples (colored bars indicate R-loop peaks). **C.** Bar chart of DRIP-qPCR (as fold change over control) for *DGCR2* using indicated qPCR primer pairs. Each bar is the average of three-independent experiments (shown with SE). **D.** Genome browser screenshot showing the same region as in (B) but now displaying EU-seq signals. **E.** Metaplot of DRIP-seq signals over gene body (as % of gene length) and terminal regions (+/- 5kb from PAS) of EL genes. Control and PladB-treated samples are color-coded as indicated. For each time point, the signal is shown as a trimmed mean (line) surrounded by the standard error (shaded). **F.** Same as E except EU-seq signals are plotted.

FIGURE 5: A. Metaplot of EU-seq signals over DoG, unaffected, and EL gene promoters (+/- 5 kb from TSS) under control and PladB-treated conditions. The relative peak increases

measured at the promoter-proximal peak and 2 kb downstream of TSS are indicated. **B**. Ratio of mean EU-seq signals (PladB/control) for DoG, unaffected, and EL genes plotted over gene body for control and PladB-treated conditions. The positions at which PladB nascent transcription signals fall below control signals are highlighted by vertical arrows. **C.** Events of intronic read count gains were identified along EL genes and categorized as intron retention (IR), alternative 3'-UTR or alternative 5'-UTR (see inset). The relative distribution of these events among these categories is plotted. **D.** The distribution of intron retention events along DoG, unaffected, and EL genes is plotted along gene bodies for the 4 hours PladB samples. The dashed lines indicate the median position for each sample class with x-intercepts indicated by numbers. P-values were determined by a Wilcoxon Mann-Whitney test. **E.** Metaplot of DRIP-seq signal under control conditions are plotted over gene bodies for DoG, unaffected, and EL genes.

FIGURE 6: A. XY plot of normalized RNA-seq counts for control and treated (4 hours PladB) samples. Each dot on the graph corresponds to a gene. Red/blue colors indicate gains/losses of RNA-seq signals (adjusted p-value of ≤ 0.05 and $|\log_2(FC)| \geq 1$). **B.** Enriched gene ontologies for RNA-seq gain genes. **C.** Enriched gene ontologies for RNA-seq loss genes. The dashed line indicates 5% FDR. **D.** Genome browser screenshot over the *RECQL5* gene showing minus strand poly(A) RNA-seq signal obtained from controls and PladB treated K562 cells. One out of three independent replicates is shown.

Figure 7: **A.** Representative immunofluorescence microscopy images from DMSO- and PladB-treated (4 and 24 hours), and Etoposide-treated (ETO) HeLa cells analyzed for γH2AX distribution and DAPI. **B**. Quantification of the immunofluorescence data for each condition; the

number of cells across 2 biological replicates analyzed per condition is indicated. **C**. γH2AX ChIP-qPCR data expressed as % input over a range of loci, treated or not with PladB, as indicated. IgG was used as a non-specific negative control on the same samples. The DIvA cell line (shown at right) in which site-specific breaks can be induced was used as a positive control for the experiment. The induction of γH2AX recruitment in DIvA cells was statistically significant ($p<0.05$). **D.** Model – see text for details.

Acknowledgements

Work in the Chedin lab is supported by the National Institutes of Health grand R01 GM120607 (to F.C). D.C.G. was supported in part by the NIH T32 predoctoctoral training program in Molecular and Cellular Biology (GM007377) and by an NIH F31 individual fellowship (GM136143). The DIvA cell line was a kind gift from Dr. Giovanni Tonon (San Raffaele Scientific Institute, Milan, Italy). Hydra RNA was a kind gift from Dr. Celina Juliano. We thank members of the Chedin lab for useful suggestions.

Authors contributions

D.C.G performed all experiments with the assistance of L.A.S for R-loop mapping. S.R.H. performed computational data analysis with help from D.C.G. F.C. conceived and supervised the project, analyzed data, and wrote the manuscript along with all authors.

Declaration of Interests

The authors report no competing interests.

Data availability

All high-throughput genome sequencing datasets are available on the NCBI GEO website under accession number GSE148768.

STAR* Methods

Analysis of transcriptome alterations by RNA-seq and RT-qPCR

Human K562 cells were treated with PladB (100 nM final concentration) for 2 and 4 hours, at which point they were harvested. Mock-treated (DMSO) and untreated samples were also collected as controls. At each time point, total RNA was isolated from $1x10_6$ cells using TRI reagent (Life Technologies, Grand Island, NY) and Direct-zol RNA MiniPrep kit (Zymo Research). Isolated RNA was then treated with 2.5 µl of DNase I (NEB) for 1 hour at 37°C. Following DNase I treatment, RNA was cleaned up using RNA Clean & Concentrator-5 Kit (Zymo Research) and resuspended in 15 µl. After cleanup, poly(A) RNA was selected and RNA-seq libraries were constructed using the Illumina TruSeq kit prior to sequencing on an Illumina HiSeq 4000 instrument. Mapping and all downstream analysis were performed on two independent biological replicates.

For RT-qPCR, cDNA was generated using iScript Reverse Transcription Supermix for RT-qPCR (Bio-Rad), with 900 ng of human RNA and 100 ng of Hydra RNA (as exogenous normalizer). cDNA was diluted 1:3 (final volume 60 µl). For RT-qPCR, 1 µl of 1:3 dilution of the cDNA was used per well in a 20 µl reaction, along with 2 µl of 10 µM primer sets, and 10 µl of SsoAdvanced Universal SYBR Green Supermix (Bio-Rad). Reactions were run in triplicate on a CFX96 Touch Real-Time PCR Detection System (Bio-Rad) with the following protocol: 95°C (30 sec), 40 cycles of 95°C (10 sec) and 95°C (30 sec), followed by melt curve analysis (65-95°C). Quantification was calculated using the $\Delta\Delta C_t$ method. All values were normalized to a Hydra gene (RP49) and to time zero (Untreated/DMSO). RT-qPCR was performed on three independent biological replicates.

Analysis of R-loop genomic distribution by DRIP-seq and DRIP-qPCR

PladB treated and control K562 cells were harvested 2 and 4 hours after PladB treatment. At each time point, genomic DNA was isolated from $5x10_6$ cells and extracted as described (Sanz and Chedin, 2019). S9.6-based DNA:RNA hybrid immunoprecipitation (DRIP) was performed as described previously (Ginno et al., 2012). In brief, 4.4 µg of restriction digested genomic DNA was immunoprecipitated for 16 hours at 4°C with 10 µL of S9.6 antibody. Three immunoprecipitated samples were combined into one for library construction. Once the quality of the combined immunoprecipitation was checked by qPCR at a range of positive and negative test loci, barcoded sequencing libraries were built, library quality was verified, and sequencing was performed on Illumina HiSeq or NovaSeq instruments. Mapping and all downstream analysis was performed on two independent biological replicates. R-loops were also mapped using sDRIP-seq, a variant of DRIP-seq that permits strand-specific R-loop mapping after genome fragmentation using sonication (Smolka et al., 2020).

For DRIP-qPCR, 2 µl of immunoprecipitated DNA (and 1:10 diluted input DNA) was used per well in a 20 µl reaction, along with 2 µl of 10 µM primer sets, and 10 µl of SsoAdvanced Universal SYBR Green Supermix (Bio-Rad). Reactions were run in duplicate on a CFX96 Touch Real-Time PCR Detection System (Bio-Rad) with the following protocol: 95°C (30 sec), 40 cycles of 95°C (10 sec) and 95°C (30 sec), followed by melt curve analysis (65-95°C). All values were normalized relative to input and normalized to Control (Untreated & DMSO) to calculate fold enrichment. DRIP-qPCR was performed on three independent biological replicates.

Analysis of nascent transcription using EU-seq

Human K562 cells treated or not with PladB were incubated with 5-ethynyl uridine (EU) (0.5 mM final concentration) for 20 min prior to harvesting two- and four-hours post PladB treatment. At each time point total RNA was isolated from $3x10_6$ cells and initially processed as described for RNA-seq. EU-RNA was enriched using Click-iT Nascent RNA Capture Kit (Invitrogen), with 5 µg of EU-RNA for biotinylation by click reaction and 1 µg of biotinylated RNA for binding to Streptavidin T1 magnetic beads. cDNA was generated using iScript Select cDNA Synthesis Kit (Bio-Rad). cDNA was cleaned up before second-strand synthesis. Then cDNA was ligated to Illumina barcoded adapters for sequencing. Library quality was checked on an Agilent BioAnalyzer and sequencing were performed on Illumina HiSeq or NovaSeq instruments. Mapping and downstream analysis was performed on two independent biological replicates.

Sequencing data processing and computational data analysis.

STAR version 2.7 (2.7.0f_0328) was used to map all sequencing reads and splice junctions (default parameters). Macs2 v2.2.5 was used to call EU-seq and sDRIP-seq signals according to defined bins. For DRIP-seq, we used restriction digest as bins. For RNA-seq, we used annotated genes and gene features (e.g. exon, intron) as their bins using Gencode version 19 (GRCh37). For subsequent analysis, the gene set was further filtered for highest APPRIS principal protein-coding genes with expression levels of at least 10 RPKM, as an annotation file. Htseq-count v0.11.2 was then used to count reads into their bins. DESeq2 v1.24 was used to calculate significant gains or losses (adjusted p-value of ≤ 0.05 and |log2(FC)| ≥ 1). rMATS v4.0.2 were used to identify significant changes (adjusted p-value of ≤ 0.05 and |log2(FC)| ≥ 0.32) in skipped exons (SE), A3SS, A5SS, and intron retentions (IR). In addition to rMATS, intron retention events were called by identifying significant increases in intronic reads using DESeq2 as well as splice junctions from STAR. To distinguish IR events from other possible events, intronic gains that spanned less than 75% of intron length were categorized into putative

3'UTRs if their 5' end were adjoined (within +/- 10% of intron length) to the 3' end of an exon or junction, and their 3' end were not adjoined to the 5' end of an exon or junction (and vice versa for putative 5'UTRs). For location analysis, we divided the genic space into several regions: promoters (+/-1kb of TSS), gene body (+1kb of TSS to -1kb of PAS), terminal (+/- 1kb of PAS), downstream of genes (DoG), and intergenic regions (everything else). DoG regions were defined as regions between a gene's terminal region (+1kb of PAS + 1 kb) up to either 150 kb downstream if no neighbor was found prior, or to the promoter of its closest downstream gene minus 1 kb, whichever is closest. The 150 kb limit was chosen because a few test genes showed DoG regions extending more than 100 kb. Post-analysis, very few genes had DoG regions longer than 100 kb. To correctly assign a DoG region to the gene from which it originated, we annotated the strandedness of EU-seq and DRIP-seq signals based on the ratio of positive and negative strands of RNA-seq and sDRIP-seq signals.

Fixation and labelling for γH2AX immunofluorescence

HeLa cells were grown, fixed, permeabilized, washed, immunostained, and imaged in 35 mm glass bottom poly-D-lysine-coated dishes (P35GC-1.5-14-C, MatTek) using 2 mL volumes of media and buffer solutions. All steps for fixation and immunofluorescence were carried out at room temperature. Cells were fixed in freshly prepared 1% formaldehyde in PBS for 10 minutes. Fixation solutions were quenched with the addition of 200 µL of 1 M glycine in PBS. Samples were washed once with PBS, and then incubated in permeabilization buffer (PBS with 0.1% Triton X-100) for 10 minutes. Samples were then incubated in staining buffer (TBST with 0.1% BSA (A9647-50G, Sigma)) for 10 minutes with rocking and a 1:1000 dilution of primary anti-phospho-Histone H2A.X (Ser139) antibody (05-636, Sigma) was added for 1 hour with rocking. Samples were then washed once with staining buffer and incubated with a 1:2000 dilution of secondary anti-mouse AlexaFluor 488 conjugate (A28175, Invitrogen) for 1 hour with rocking

with samples kept concealed from light from this step onward. A 2.5 µg/mL DAPI solution in staining buffer was then added for 2 minutes and washed in TBST for 10 minutes with rocking. Samples were then immediately imaged. For each experiment, all samples were prepared, treated, and stained in parallel from one master antibody and/or dye dilution to ensure uniform treatment and staining efficiencies. Fixation and labelling was performed for two independent biological replicates.

Imaging and image analysis

Fixed cell imaging was performed using a confocal laser scanning biological microscope [Olympus FV1000] with a 60X objective. For each experiment, all conditions were imaged in parallel with identical exposure times and laser settings. Images were analyzed and quantified using ImageJ. Statistics and data visualization were done using RStudio. P-values were determined by a Wilcoxon Mann-Whitney test using wilcox.test() function. The number of foci were counted for individual nuclear regions defined by thresholding the DAPI channel to obtain Regions of Interest (ROI) for analysis. ROIs containing more than one nucleus were removed from downstream analysis. Find Maxima was applied to the 488 channel using a noise tolerance of 700 to generate a binary image of local maxima (γH2AX foci). Analysis was then performed on nuclear ROIs to quantify the number of γH2AX foci per nucleus.

Analysis of γH2AX genomic distribution using ChIP-qPCR

Human K562 cells ($1x10_7$ cells) were treated with PladB (100 nM final concentration) for 4 and 24 hours or DMSO as mock-control. DIvA cells were treated with or without 4-hydroxy tamoxifen (4OHT) (300 nM final concentration) for 4 hours. K562 cells were then collected and resuspended in PBS, then formaldehyde was added to a final concentration of 1% and

crosslinking was allowed to proceed for 10 minutes at room temperature, while the samples were rocking. For DIvA cells, media was removed and replaced with PBS and Formaldehyde was added to dish. In order to stop the reaction, fresh glycine was added to a final concentration of 0.125 M. For DIvA cells, the cells were then scraped and washed once with PBS. After 10 minutes cells were washed once with PBS. Pelleted cells were then resuspended in RIPA lysis buffer (150 mM NaCl, 5 mM EDTA, 50 mM Tris pH 8.0, 1% IGEPAL CA-630, 0.5% sodium deoxycholate, 0.1% SDS) supplemented with phosphatase and protease inhibitors (PHOSS-RO, Sigma and 539134, Millipore). The lysate was then sonicated 30 secs for 15 cycles (Diagenode Bioruptor) to obtain DNA fragments of about 500-1000 bp. Samples were then centrifuged for 15 mins at 4ºC at 15,000 rpm and the supernatants were transferred to new tubes. Samples were then diluted in RIPA buffer (100 µl chromatin : 900 µl RIPA buffer) and phosphatase and protease inhibitors were added to samples. A portion of the sample was removed and saved as Input. The rest of the sample was subjected to a 60 min preclearing with 60 µl of previously blocked protein A/G magnetic agarose beads (PI78609, Thermo Scientific) at 4ºC. Blocking was achieved by incubating the agarose beads in RIPA buffer with 0.1% BSA for 30 minutes. Precleared samples were incubated with rotation overnight at 4ºC with either 10 µg of anti-phospho-Histone H2A.X (Ser139) antibody (05-636, Sigma) or 10 µg of IgG as control. Immune complexes were then recovered by incubating the samples with 60 µl of previously blocked protein A/G magnetic agarose beads for 1 hour at 4ºC on a rotating wheel. Beads were washed 6 times in wash buffer (150 mM NaCl, 10 mM Tris pH 7.5, 2 mM EDTA, 0.1% IGEPAL CA-630) for 10 minutes followed by an additional wash with 10 mM Tris-EDTA pH 8.0. Elution and crosslink reversal were achieved by incubating samples in 10 mM Tris-EDTA + 0.1% SDS and proteinase K at 65ºC for 4 hours. Elutions were then transferred to a phase-lock separation tube. The beads were then washed one more time with 10 mM Tris-EDTA and 0.5 M NaCl for 10 minutes at 65ºC and the supernatant was added to the phase-lock separation tube. DNA was extracted by the phenol-chloroform method and precipitated with ethanol.

For ChIP-qPCR, 2 µl of immunoprecipitated DNA (and 1:4.5 diluted input DNA) was used per well in a 20 µl reaction, along with 2 µl of 10 µM primer sets, and 10 µl of SsoAdvanced Universal SYBR Green Supermix (Bio-Rad). Reactions were run in duplicate on a CFX96 Touch Real-Time PCR Detection System (Bio-Rad) with the following protocol: 95°C (30 sec), 40 cycles of 95°C (10 sec) and 95°C (30 sec), followed by melt curve analysis (65-95°C). All values were normalized relative to input. ChIP-qPCR was performed on two independent biological replicates for K562 samples and two technical replicates for DIvA samples.

SUPPLEMENTAL FIGURE LEGENDS

Figure S1: A. Gel-based RT-PCR splicing assay verifies that PladB treatment causes intron retention over the *DNAJB1* target gene, but not the control *ACTB* gene. No PCR products were obtained if reverse-transcription (RT) was omitted, as expected. **B.** RT-qPCR assay confirms that PladB treatment causes intron retention for *DNAJB1* target gene. IR starts to occur as early as 15 min. **C.** RT-qPCR assay confirms that PladB treatment causes intron retention for *RPL13A*, a positive R-loop forming gene. **D.** Bar chart of DRIP-qPCR (as percent input) for *RPL13A*. Each bar is the average of three-independent experiments (shown with SE). PladB treatment results in loss of R-loops over intron with IR.

Figure S2: A. Genome browser screenshot over the *VAPA* gene and ~200 kb downstream showing plus and minus strand sDRIP-seq signal obtained from controls and PladB-treated K562 samples. R-loop gains occur directly downstream of gene (DoG) only on the plus strand (colored bar). **B.** Genome browser screenshot over the *CYCS* gene and downstream region showing DRIP-seq signal (Top) and plus and minus strand sDRIP-seq signal (Bottom) obtained from controls and PladB treated K562 samples. R-loop gains occur directly downstream of gene (DoG) and only on minus strand (colored bar). **C.** Genome browser screenshot over the *RPL9* gene and downstream region showing DRIP-seq signal (Top) and plus and minus strand sDRIP-seq signal (Bottom) obtained from controls and PladB treated K562 samples. R-loop gains occur directly downstream of gene (DoG) and only on minus strand (colored bar). **D.** Bar chart of DRIP-qPCR (as fold change over control) for a region downstream of *CYCS* PAS. Each bar is the average of three-independent experiments (shown with SE). **E.** Bar chart of DRIP-qPCR (as fold change over control) for *RPL9* and a region downstream of *RPL9* PAS. Each bar is the average of three-independent experiments (shown with SE). **F.** Genome browser

screenshot showing the same region as in (B) but now displaying EU-seq signals. Region with increased R-loops also have increased nascent transcription. **G.** Genome browser screenshot showing the same region as in C but now displaying EU-seq signals. Region with increased R-loops also have increased nascent transcription.

Figure S3: A. Location analysis of R-loop peak gains for PladB-treated K562 sample (4 hours) according to downstream of gene (DoG) annotation pipeline. Annotation scheme is below. 82% of R-loop gains 4 hours post-PladB treatment specifically map to downstream of gene (DoG) regions. **B.** STRING interaction analysis. All 429 DoG genes were used as input for STRING analysis and a network was built based on highest confidence (0.9). Nodes are color coded based on biological processes. Nodes that had no interaction are not shown.

Figure S4: A. Genome browser screenshot over the *CHCHD3* gene showing DRIP-seq (Top) and EU-seq signals (Bottom) from controls and PladB-treated K562 samples. R-loop loss is directional with transcription (colored bar). **B.** Genome browser screenshot over the *FAM120B* gene showing DRIP-seq signal (Top) and EU-seq signal (Bottom) obtained from controls and PladB treated K562 samples. R-loop loss is directional with transcription (colored bar). **C.** Genome browser screenshot over the *VRK3* gene showing DRIP-seq signal (Top),and EU-seq (Bottom) obtained from controls and PladB treated K562 samples. **D.** Bar charts of DRIP-qPCR (as fold change over control) for *CHCHD3, FAM120B, VRK3* Each bar is the average of three-independent experiments (shown with SE). Primer locations are indicated on each panel.

Figure S5: A. Gene length distribution for DoG, unaffected, and EL genes. **B.** Distribution of distances between DoG, unaffected, and EL genes and their nearest gene neighbor upstream

or downstream. **C.** Travel ratios for DoG, EL and unaffected genes are plotted using EU-seq signals under control conditions. **D.** Travel ratios for DoG, EL and unaffected genes are plotted using RNAP II ChIP-seq signals. **E.** Genome browser screenshot over the *TP53* gene showing minus strand poly(A) RNA-seq signal obtained from control and PladB-treated K562 cells. An enrichment of poly-adenylated alternative 3'-UTR reads is clearly detectable in PladB-treated samples. **F.** The positions of the earliest DRIP-seq loss events along EL genes were annotated, and their distribution is plotted along gene bodies (4 hours post PladB samples). **G.** The distribution of skipped exon events along DoG, unaffected, and EL genes is plotted along gene bodies 4 hours post PladB treatment. The dashed lines indicate the position of the median position of such events. All p-values were determined by a Wilcoxon Mann-Whitney test (one, two, and three stars indicate p-value of ≤ 0.05, 0.01, and 0.001, respectively; NS: Not Significant).

FIGURE S6: A. STRING interaction analysis. All 721 RNA-seq gain genes were used as input for STRING analysis and a network was built based on highest confidence (0.9). Nodes are color coded based on biological processes. Nodes that had no interaction are not shown. **B.** Similar to (A) except all 1,180 RNA-seq loss genes were used as input for STRING analysis. **C.** Pie chart comparing the overlap in the number of loss events for RNA-seq, DRIP-seq, and EU-seq. **D.** Genome browser screenshot over the *RECQL5* gene showing DRIP-seq signal (Top) and EU-seq signal (Bottom) obtained from controls and PladB-treated K562 samples. R-loop loss is directional with transcription (colored bar).

REFERENCES

Adelman, K., and Lis, J.T. (2012). Promoter-proximal pausing of RNA polymerase II: emerging roles in metazoans. Nat Rev Genet *13*, 720-731.

Aguilera, A. (2005). mRNA processing and genomic instability. Nat Struct Mol Biol *12*, 737-738.

Aguilera, A., and Garcia-Muse, T. (2012). R loops: from transcription byproducts to threats to genome stability. Mol Cell *46*, 115-124.

Bauer, D.L.V., Tellier, M., Martinez-Alonso, M., Nojima, T., Proudfoot, N.J., Murphy, S., and Fodor, E. (2018). Influenza Virus Mounts a Two-Pronged Attack on Host RNA Polymerase II Transcription. Cell Rep *23*, 2119-2129 e2113.

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

Berg, M.G., Singh, L.N., Younis, I., Liu, Q., Pinto, A.M., Kaida, D., Zhang, Z., Cho, S., Sherrill-Mix, S., Wan, L.*, et al.* (2012). U1 snRNP determines mRNA length and regulates isoform expression. Cell *150*, 53-64.

Blasco, M.A. (2007). The epigenetic regulation of mammalian telomeres. Nat Rev Genet *8*, 299-309.

Bonnet, A., Grosso, A.R., Elkaoutari, A., Coleno, E., Presle, A., Sridhara, S.C., Janbon, G., Geli, V., de Almeida, S.F., and Palancade, B. (2017). Introns Protect Eukaryotic Genomes from Transcription-Associated Genetic Instability. Mol Cell *67*, 608-621 e606.

Boswell, S.A., Snavely, A., Landry, H.M., Churchman, L.S., Gray, J.M., and Springer, M. (2017). Total RNA-seq to identify pharmacological effects on specific stages of mRNA synthesis. Nat Chem Biol *13*, 501-507.

Cardiello, J.F., Goodrich, J.A., and Kugel, J.F. (2018). Heat Shock Causes a Reversible Increase in RNA Polymerase II Occupancy Downstream of mRNA Genes, Consistent with a Global Loss in Transcriptional Termination. Mol Cell Biol *38*.

Carvalho, T., Martins, S., Rino, J., Marinho, S., and Carmo-Fonseca, M. (2017). Pharmacological inhibition of the spliceosome subunit SF3b triggers exon junction complex-independent nonsense-mediated decay. J Cell Sci *130*, 1519-1531.

Cerritelli, S.M., and Crouch, R.J. (2009). Ribonuclease H: the enzymes in eukaryotes. Febs J *276*, 1494-1505.

Chan, Y.A., Hieter, P., and Stirling, P.C. (2014). Mechanisms of genome instability induced by RNA-processing defects. Trends Genet *30*, 245-253.

Chedin, F. (2016). Nascent Connections: R-Loops and Chromatin Patterning. Trends Genet *32*, 828-838.

Chedin, F., and Benham, C.J. (2020). Emerging roles for R-loop structures in the management of topological stress. J Biol Chem *295*, 4684-4695.

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

Cretu, C., Agrawal, A.A., Cook, A., Will, C.L., Fekkes, P., Smith, P.G., Luhrmann, R., Larsen, N., Buonamici, S., and Pena, V. (2018). Structural Basis of Splicing Modulation by Antitumor Macrolide Compounds. Mol Cell *70*, 265-273 e268.

Crossley, M.P., Bocek, M., and Cimprich, K.A. (2019). R-Loops as Cellular Regulators and Genomic Threats. Mol Cell *73*, 398-411.

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.

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 e988.

Drolet, M., Broccoli, S., Rallu, F., Hraiky, C., Fortin, C., Masse, E., and Baaklini, I. (2003). The problem of hypernegative supercoiling and R-loop formation in transcription. Front Biosci *8*, d210-221.

Effenberger, K.A., Anderson, D.D., Bray, W.M., Prichard, B.E., Ma, N., Adams, M.S., Ghosh, A.K., and Jurica, M.S. (2014). Coherence between cellular responses and in vitro splicing inhibition for the anti-tumor drug pladienolide B and its analogs. J Biol Chem *289*, 1938-1947.

Garcia-Muse, T., and Aguilera, A. (2019). R Loops: From Physiological to Pathological Roles. Cell *179*, 604-618.

Giannakakis, A., Zhang, J., Jenjaroenpun, P., Nama, S., Zainolabidin, N., Aau, M.Y., Yarmishyn, A.A., Vaz, C., Ivshina, A.V., Grinchuk, O.V.*, et al.* (2015). Contrasting expression patterns of coding and noncoding parts of the human genome upon oxidative stress. Sci Rep *5*, 9737.

Ginno, P.A., Lott, P.L., Christensen, H.C., Korf, I., and Chedin, F. (2012). R-loop formation is a distinctive characteristic of unmethylated human CpG island promoters. Mol Cell *45*, 814-825.

Gressel, S., Schwalb, B., and Cramer, P. (2019). The pause-initiation limit restricts transcription activation in human cells. Nat Commun *10*, 3603.

Gressel, S., Schwalb, B., Decker, T.M., Qin, W., Leonhardt, H., Eick, D., and Cramer, P. (2017). CDK9-dependent RNA polymerase II pausing controls transcription initiation. Elife *6*.

Hamperl, S., and Cimprich, K.A. (2014). The contribution of co-transcriptional RNA:DNA hybrid structures to DNA damage and genome instability. DNA Repair (Amst) *19*, 84-94.

Hartono, S.R., Malapert, A., Legros, P., Bernard, P., Chedin, F., and Vanoosthuyse, V. (2018). The Affinity of the S9.6 Antibody for Double-Stranded RNAs Impacts the Accurate Mapping of R-Loops in Fission Yeast. J Mol Biol *430*, 272-284.

Heinz, S., Texari, L., Hayes, M.G.B., Urbanowski, M., Chang, M.W., Givarkes, N., Rialdi, A., White, K.M., Albrecht, R.A., Pache, L.*, et al.* (2018). Transcription Elongation Can Affect Genome 3D Structure. Cell *174*, 1522-1536 e1522.

Hennig, T., Michalski, M., Rutkowski, A.J., Djakovic, L., Whisnant, A.W., Friedl, M.S., Jha, B.A., Baptista, M.A.P., L'Hernault, A., Erhard, F.*, et al.* (2018). HSV-1-induced disruption of transcription termination resembles a cellular stress response but selectively increases chromatin accessibility downstream of genes. PLoS Pathog *14*, e1006954.

Herzel, L., Ottoz, D.S.M., Alpert, T., and Neugebauer, K.M. (2017). Splicing and transcription touch base: co-transcriptional spliceosome assembly and function. Nat Rev Mol Cell Biol *18*, 637-650.

Hsin, J.P., and Manley, J.L. (2012). The RNA polymerase II CTD coordinates transcription and RNA processing. Genes Dev *26*, 2119-2137.

Huertas, P., and Aguilera, A. (2003). Cotranscriptionally formed DNA:RNA hybrids mediate transcription elongation impairment and transcription-associated recombination. Mol Cell *12*, 711-721.

Iacovoni, J.S., Caron, P., Lassadi, I., Nicolas, E., Massip, L., Trouche, D., and Legube, G. (2010). High-resolution profiling of gammaH2AX around DNA double strand breaks in the mammalian genome. EMBO J *29*, 1446-1457.

Kaida, D., Berg, M.G., Younis, I., Kasim, M., Singh, L.N., Wan, L., and Dreyfuss, G. (2010). U1 snRNP protects pre-mRNAs from premature cleavage and polyadenylation. Nature *468*, 664-668.

Kashyap, M.K., Kumar, D., Villa, R., La Clair, J.J., Benner, C., Sasik, R., Jones, H., Ghia, E.M., Rassenti, L.Z., Kipps, T.J.*, et al.* (2015). Targeting the spliceosome in chronic lymphocytic leukemia with the macrolides FD-895 and pladienolide-B. Haematologica *100*, 945-954.

Kim Guisbert, K.S., and Guisbert, E. (2017). SF3B1 is a stress-sensitive splicing factor that regulates both HSF1 concentration and activity. PLoS One *12*, e0176382.

Koga, M., Hayashi, M., and Kaida, D. (2015). Splicing inhibition decreases phosphorylation level of Ser2 in Pol II CTD. Nucleic Acids Res *43*, 8258-8267.

Kotake, Y., Sagane, K., Owa, T., Mimori-Kiyosue, Y., Shimizu, H., Uesugi, M., Ishihama, Y., Iwata, M., and Mizui, Y. (2007). Splicing factor SF3b as a target of the antitumor natural product pladienolide. Nat Chem Biol *3*, 570-575.

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.

Li, X., and Manley, J.L. (2005). Inactivation of the SR protein splicing factor ASF/SF2 results in genomic instability. Cell *122*, 365-378.

Li, X., and Manley, J.L. (2006). Cotranscriptional processes and their influence on genome stability. Genes Dev *20*, 1838-1847.

Mahat, D.B., Salamanca, H.H., Duarte, F.M., Danko, C.G., and Lis, J.T. (2016). Mammalian Heat Shock Response and Mechanisms Underlying Its Genome-wide Transcriptional Regulation. Mol Cell *62*, 63-78.

Malig, M., Hartono, S.R., Giafaglione, J.M., Sanz, L.A., and Chedin, F. (2020). Ultra-Deep Coverage Single-Molecule R-loop Footprinting Reveals Principles of R-loop Formation. J Mol Biol.

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.

Metkar, M., Ozadam, H., Lajoie, B.R., Imakaev, M., Mirny, L.A., Dekker, J., and Moore, M.J. (2018). Higher-Order Organization Principles of Pre-translational mRNPs. Mol Cell *72*, 715-726 e713.

Nguyen, H.D., Leong, W.Y., Li, W., Reddy, P.N.G., Sullivan, J.D., Walter, M.J., Zou, L., and Graubert, T.A. (2018). Spliceosome Mutations Induce R Loop-Associated Sensitivity to ATR Inhibition in Myelodysplastic Syndromes. Cancer Res *78*, 5363-5374.

Nguyen, H.D., Yadav, T., Giri, S., Saez, B., Graubert, T.A., and Zou, L. (2017). Functions of Replication Protein A as a Sensor of R Loops and a Regulator of RNaseH1. Mol Cell *65*, 832-847 e834.

Nguyen, H.D., Zou, L., and Graubert, T.A. (2019). Targeting R-loop-associated ATR response in myelodysplastic syndrome. Oncotarget *10*, 2581-2582.

Niehrs, C., and Luke, B. (2020). Regulatory R-loops as facilitators of gene expression and genome stability. Nat Rev Mol Cell Biol *21*, 167-178.

Nojima, T., Gomes, T., Grosso, A.R., 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.

O'Brien, K., Matlin, A.J., Lowell, A.M., and Moore, M.J. (2008). The biflavonoid isoginkgetin is a general inhibitor of Pre-mRNA splicing. J Biol Chem *283*, 33147-33154.

Oh, J.M., Di, C., Venters, C.C., Guo, J., Arai, C., So, B.R., Pinto, A.M., Zhang, Z., Wan, L., Younis, I.*, et al.* (2017). U1 snRNP telescripting regulates a size-function-stratified human genome. Nat Struct Mol Biol *24*, 993-999.

Paulsen, R.D., Soni, D.V., Wollman, R., Hahn, A.T., Yee, M.C., Guan, A., Hesley, J.A., Miller, S.C., Cromwell, E.F., Solow-Cordero, D.E.*, et al.* (2009). A genome-wide siRNA screen reveals diverse cellular processes and pathways that mediate genome stability. Mol Cell *35*, 228-239.

Phillips, D.D., Garboczi, D.N., Singh, K., Hu, Z., Leppla, S.H., and Leysath, C.E. (2013). The sub-nanomolar binding of DNA-RNA hybrids by the single-chain Fv fragment of antibody S9.6. J Mol Recognit *26*, 376-381.

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

Richard, P., and Manley, J.L. (2017). R Loops and Links to Human Disease. J Mol Biol *429*, 3168-3180.

Rigo, F., and Martinson, H.G. (2008). Functional coupling of last-intron splicing and 3'-end processing to transcription in vitro: the poly(A) signal couples to splicing before committing to cleavage. Mol Cell Biol *28*, 849-862.

Rigo, F., and Martinson, H.G. (2009). Polyadenylation releases mRNA from RNA polymerase II in a process that is licensed by splicing. RNA *15*, 823-836.

Santos-Pereira, J.M., and Aguilera, A. (2015). R loops: new modulators of genome dynamics and function. Nat Rev Genet *16*, 583-597.

Sanz, L.A., and Chedin, F. (2019). High-resolution, strand-specific R-loop mapping via S9.6-based DNA-RNA immunoprecipitation and high-throughput sequencing. Nat Protoc.

Sanz, L.A., Hartono, S.R., Lim, Y.W., Steyaert, S., Rajpurkar, A., Ginno, P.A., Xu, X., and Chedin, F. (2016). Prevalent, Dynamic, and Conserved R-Loop Structures Associate with Specific Epigenomic Signatures in Mammals. Mol Cell *63*, 167-178.

Saponaro, M., Kantidakis, T., Mitter, R., Kelly, G.P., Heron, M., Williams, H., Soding, J., Stewart, A., and Svejstrup, J.Q. (2014). RECQL5 controls transcript elongation and suppresses genome instability associated with transcription stress. Cell *157*, 1037-1049.

Skourti-Stathaki, K., Proudfoot, N.J., and Gromak, N. (2011). Human senataxin resolves RNA/DNA hybrids formed at transcriptional pause sites to promote Xrn2-dependent termination. Mol Cell *42*, 794-805.

Skourti-Stathaki, K., Torlai Triglia, E., Warburton, M., Voigt, P., Bird, A., and Pombo, A. (2019). R-Loops Enhance Polycomb Repression at a Subset of Developmental Regulator Genes. Mol Cell *73*, 930-945 e934.

Smolka, J., Hartono, S.R., Sanz, L.A., and Chedin, F. (2020). Recognition of cellular RNAs by the S9.6 antibody creates pervasive artefacts when imaging RNA:DNA hybrids.

Stirling, P.C., Chan, Y.A., Minaker, S.W., Aristizabal, M.J., Barrett, I., Sipahimalani, P., Kobor, M.S., and Hieter, P. (2012). R-loop-mediated genome instability in mRNA cleavage and polyadenylation mutants. Genes Dev *26*, 163-175.

Stolz, R., Sulthana, S., Hartono, S.R., Malig, M., Benham, C.J., and Chedin, F. (2019). Interplay between DNA sequence and negative superhelicity drives R-loop structures. Proc Natl Acad Sci U S A *116*, 6260-6269.

Sun, C. (2020). The SF3b complex: splicing and beyond. Cell Mol Life Sci.

Tan-Wong, S.M., Dhir, S., and Proudfoot, N.J. (2019). R-Loops Promote Antisense Transcription across the Mammalian Genome. Mol Cell *76*, 600-616 e606.

Tresini, M., Warmerdam, D.O., Kolovos, P., Snijder, L., Vrouwe, M.G., Demmers, J.A., van, I.W.F., Grosveld, F.G., Medema, R.H., Hoeijmakers, J.H.*, et al.* (2015). The core spliceosome as target and effector of non-canonical ATM signalling. Nature *523*, 53-58.

Vilborg, A., Passarelli, M.C., Yario, T.A., Tycowski, K.T., and Steitz, J.A. (2015). Widespread Inducible Transcription Downstream of Human Genes. Mol Cell *59*, 449-461.

Vilborg, A., Sabath, N., Wiesel, Y., Nathans, J., Levy-Adam, F., Yario, T.A., Steitz, J.A., and Shalgi, R. (2017). Comparative analysis reveals genomic features of stress-induced transcriptional readthrough. Proc Natl Acad Sci U S A *114*, E8362-E8371.

Wan, Y., Zheng, X., Chen, H., Guo, Y., Jiang, H., He, X., Zhu, X., and Zheng, Y. (2015). Splicing function of mitotic regulators links R-loop-mediated DNA damage to tumor cell killing. J Cell Biol *209*, 235-246.

Yokoi, A., Kotake, Y., Takahashi, K., Kadowaki, T., Matsumoto, Y., Minoshima, Y., Sugi, N.H., Sagane, K., Hamaguchi, M., Iwata, M.*, et al.* (2011). Biological validation that SF3b is a target of the antitumor macrolide pladienolide. FEBS J *278*, 4870-4880.

Yu, K., Chedin, F., Hsieh, C.L., Wilson, T.E., and Lieber, M.R. (2003). R-loops at immunoglobulin class switch regions in the chromosomes of stimulated B cells. Nat Immunol *4*, 442-451.

Yu, K., and Lieber, M.R. (2019). Current insights into the mechanism of mammalian immunoglobulin class switch recombination. Crit Rev Biochem Mol Biol *54*, 333-351.

Figure 1

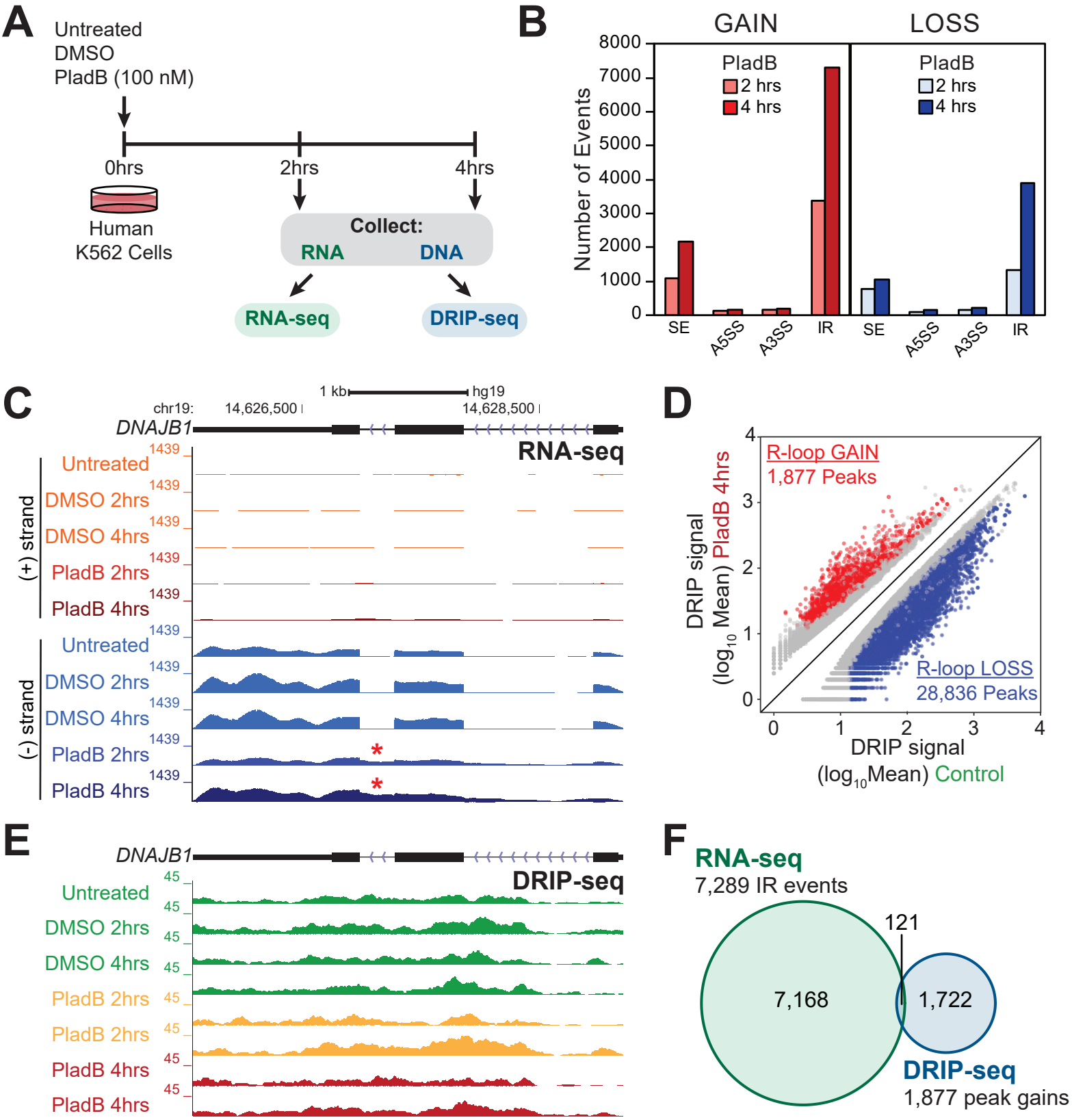

A
Untreated
DMSO
PladB (100 nM)
0hrs
2hrs
4hrs
Human
K562 Cells
Collect:
RNA
DNA
RNA-seq
DRIP-seq
B
GAIN
LOSS
Number of Events
PladB
2 hrs
4 hrs
SE
A5SS
A3SS
IR
C
1 kb
hg19
chr19:
14,626,500
14,628,500
DNAJB1
RNA-seq
(+) strand
(-) strand
Untreated
DMSO 2hrs
DMSO 4hrs
PladB 2hrs
PladB 4hrs
1439
D
R-loop GAIN
1,877 Peaks
R-loop LOSS
28,836 Peaks
DRIP signal
(log10 Mean) PladB 4hrs
DRIP signal
(log10Mean) Control
E
DNAJB1
DRIP-seq
Untreated
DMSO 2hrs
DMSO 4hrs
PladB 2hrs
PladB 2hrs
PladB 4hrs
PladB 4hrs
45
F
RNA-seq
7,289 IR events
121
7,168
1,722
DRIP-seq
1,877 peak gains

Figure 2

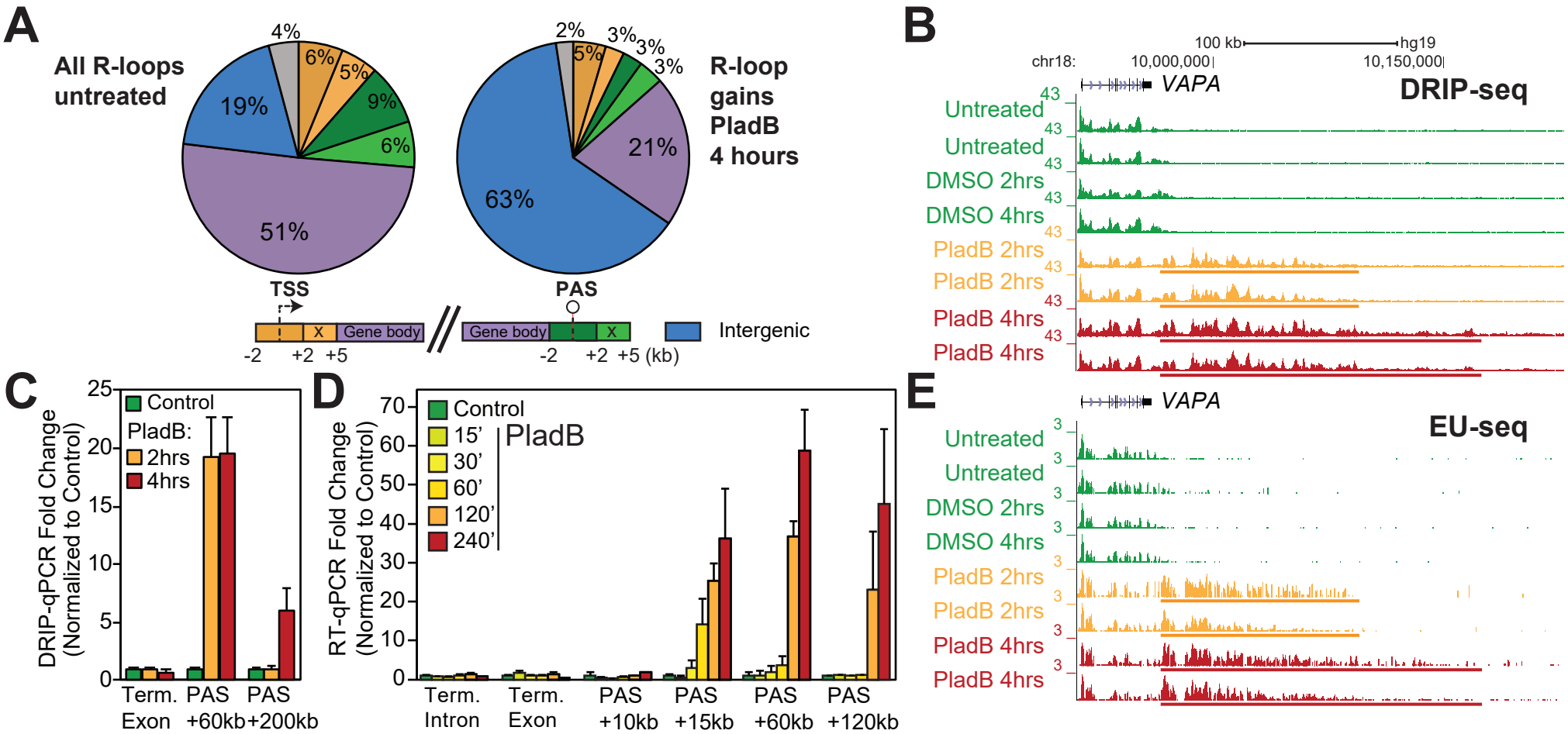

A
All R-loops untreated
4%
6%
5%
9%
6%
19%
51%
R-loop gains PladB 4 hours
2%
3%
3%
3%
5%
21%
63%
TSS
PAS
Gene body
X
-2
+2
+5
+5 (kb)
Intergenic
B
100 kb
hg19
chr18:
10,000,000
10,150,000
VAPA
DRIP-seq
Untreated
Untreated
DMSO 2hrs
DMSO 4hrs
PladB 2hrs
PladB 2hrs
PladB 4hrs
PladB 4hrs
43
C
DRIP-qPCR Fold Change (Normalized to Control)
Control
PladB:
2hrs
4hrs
Term. Exon
PAS +60kb
PAS +200kb
D
RT-qPCR Fold Change (Normalized to Control)
Control
15'
30'
60'
120'
240'
PladB
Term. Intron
Term. Exon
PAS +10kb
PAS +15kb
PAS +60kb
PAS +120kb
E
VAPA
EU-seq
Untreated
Untreated
DMSO 2hrs
DMSO 4hrs
PladB 2hrs
PladB 2hrs
PladB 4hrs
PladB 4hrs
3

Figure 3

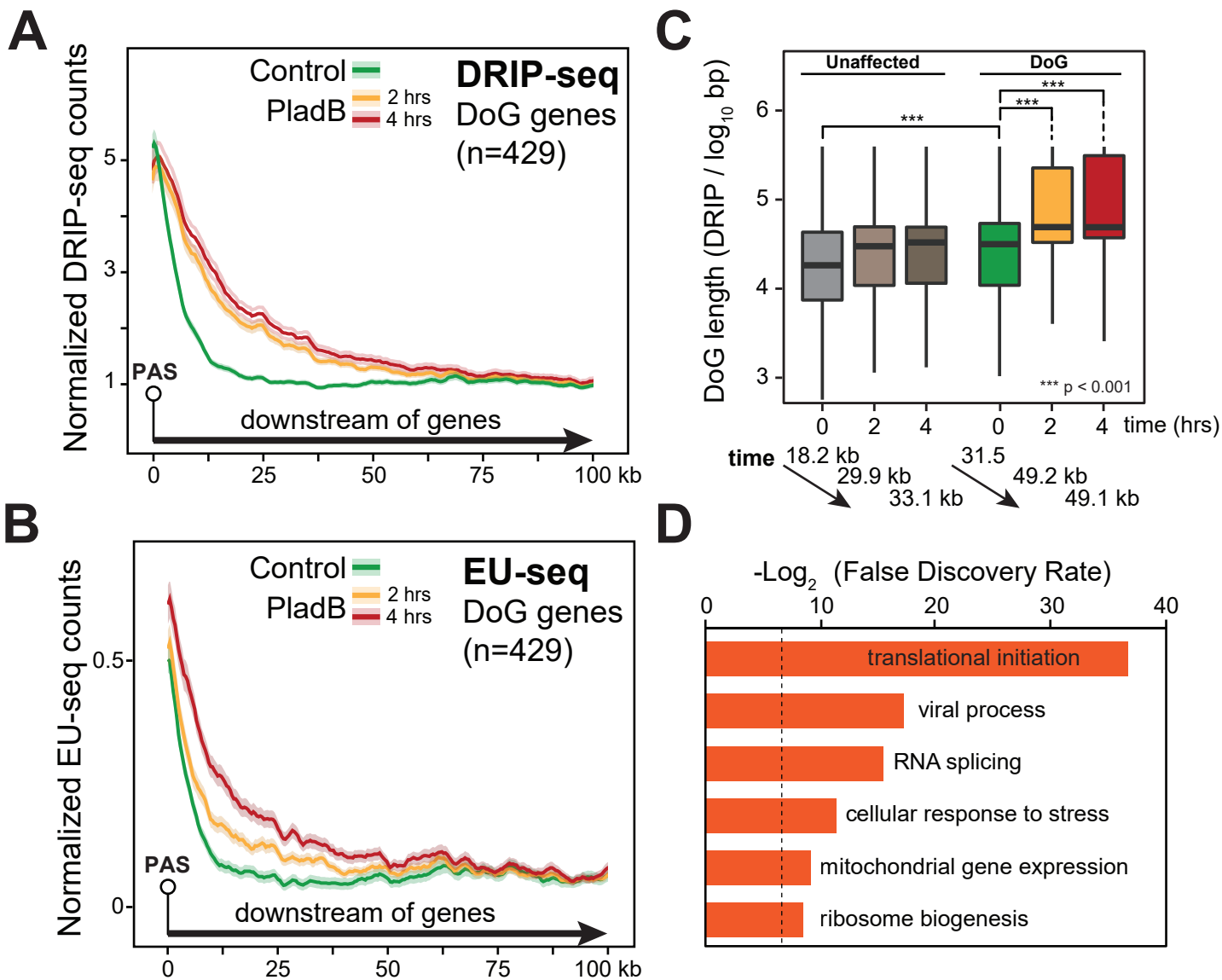

A
Control
PladB
2 hrs
4 hrs
DRIP-seq
DoG genes
(n=429)
Normalized DRIP-seq counts
PAS
downstream of genes
0
25
50
75
100 kb
B
Control
PladB
2 hrs
4 hrs
EU-seq
DoG genes
(n=429)
Normalized EU-seq counts
PAS
downstream of genes
C
Unaffected
DoG
DoG length (DRIP / log10 bp)
*** p < 0.001
time (hrs)
time
18.2 kb
29.9 kb
33.1 kb
31.5
49.2 kb
49.1 kb
D
-Log2 (False Discovery Rate)
translational initiation
viral process
RNA splicing
cellular response to stress
mitochondrial gene expression
ribosome biogenesis

Figure 4

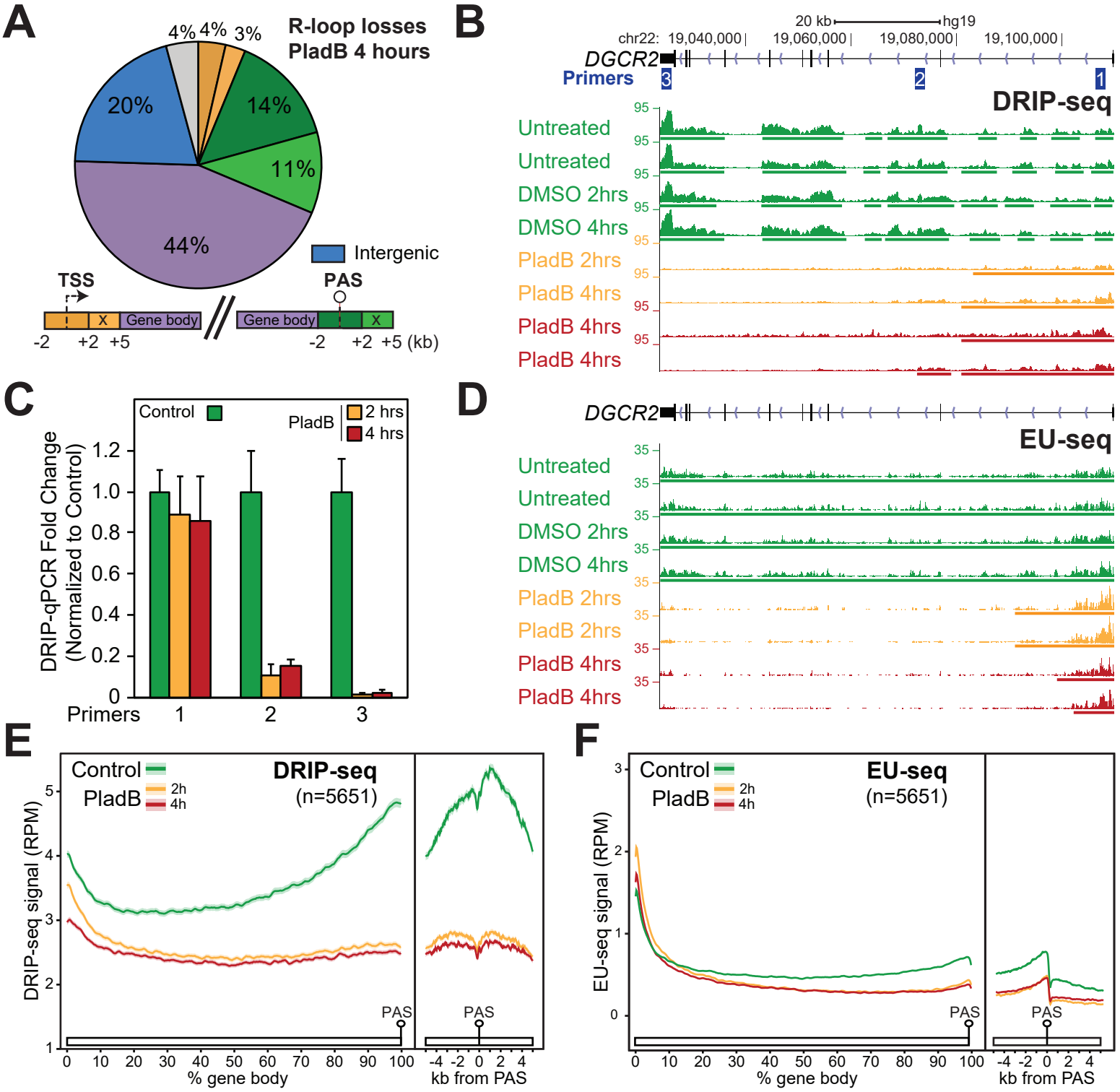

A
R-loop losses
PladB 4 hours
4%
4%
3%
14%
11%
44%
20%
Intergenic
TSS
PAS
Gene body
X
-2
+2
+5
+5 (kb)
B
20 kb
hg19
chr22: 19,040,000
19,060,000
19,080,000
19,100,000
DGCR2
Primers
DRIP-seq
Untreated
Untreated
DMSO 2hrs
DMSO 4hrs
PladB 2hrs
PladB 4hrs
PladB 4hrs
PladB 4hrs
C
Control
PladB
2 hrs
4 hrs
DRIP-qPCR Fold Change
(Normalized to Control)
Primers
1
2
3
D
DGCR2
EU-seq
Untreated
Untreated
DMSO 2hrs
DMSO 4hrs
PladB 2hrs
PladB 2hrs
PladB 4hrs
PladB 4hrs
E
Control
PladB
2h
4h
DRIP-seq
(n=5651)
DRIP-seq signal (RPM)
PAS
% gene body
kb from PAS
F
Control
PladB
2h
4h
EU-seq
(n=5651)
EU-seq signal (RPM)
PAS
% gene body
kb from PAS

Figure 5

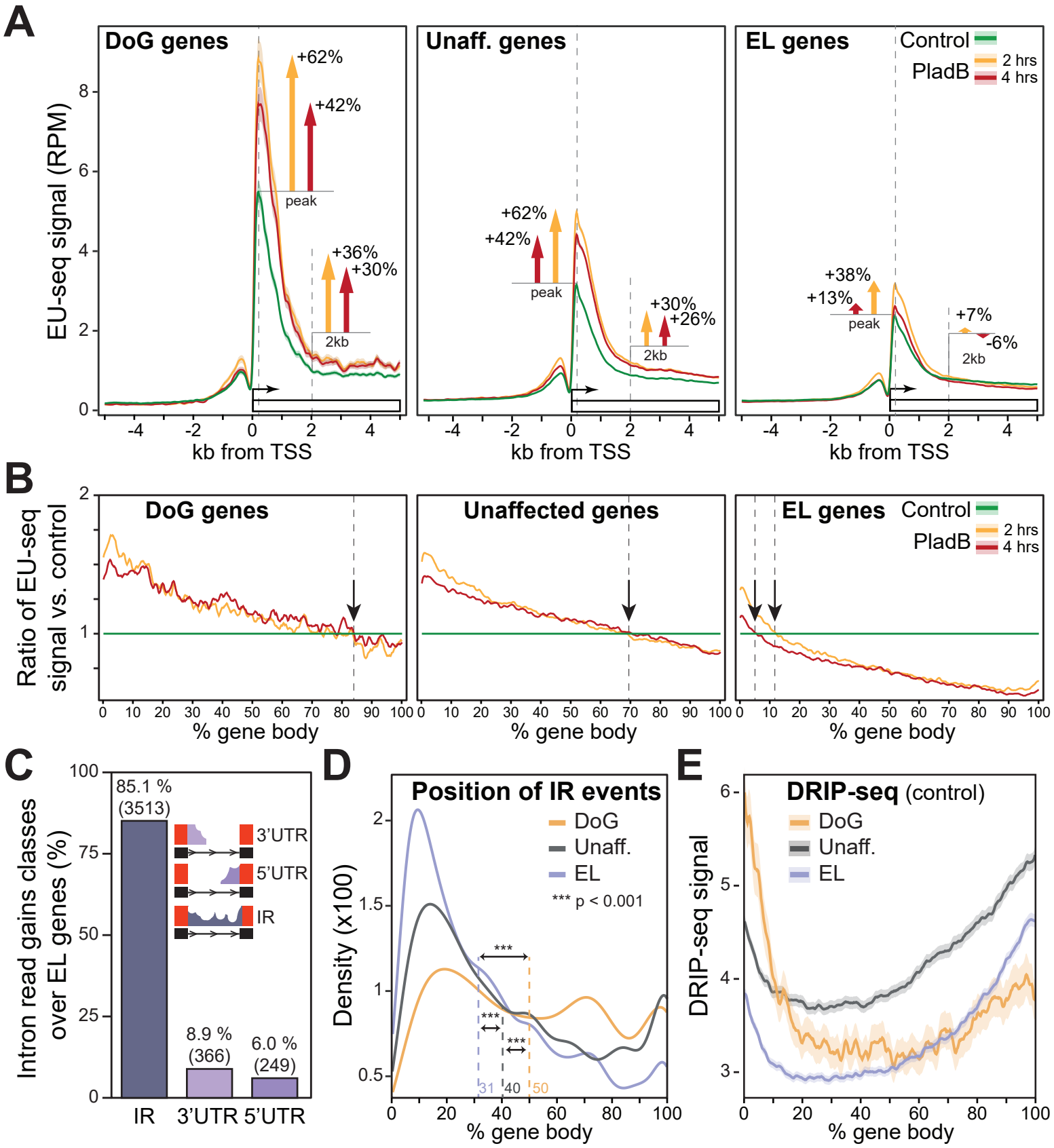

A
DoG genes
Unaff. genes
EL genes
Control
PladB
2 hrs
4 hrs
EU-seq signal (RPM)
+62%
+42%
peak
+36%
+30%
2kb
+38%
+13%
+7%
-6%
+26%
kb from TSS
B
Ratio of EU-seq signal vs. control
Unaffected genes
% gene body
C
Intron read gains classes over EL genes (%)
85.1 %
(3513)
8.9 %
(366)
6.0 %
(249)
3'UTR
5'UTR
IR
D
Position of IR events
DoG
Unaff.
EL
*** p < 0.001
Density (x100)
31
40
50
E
DRIP-seq (control)
DRIP-seq signal

Figure 6

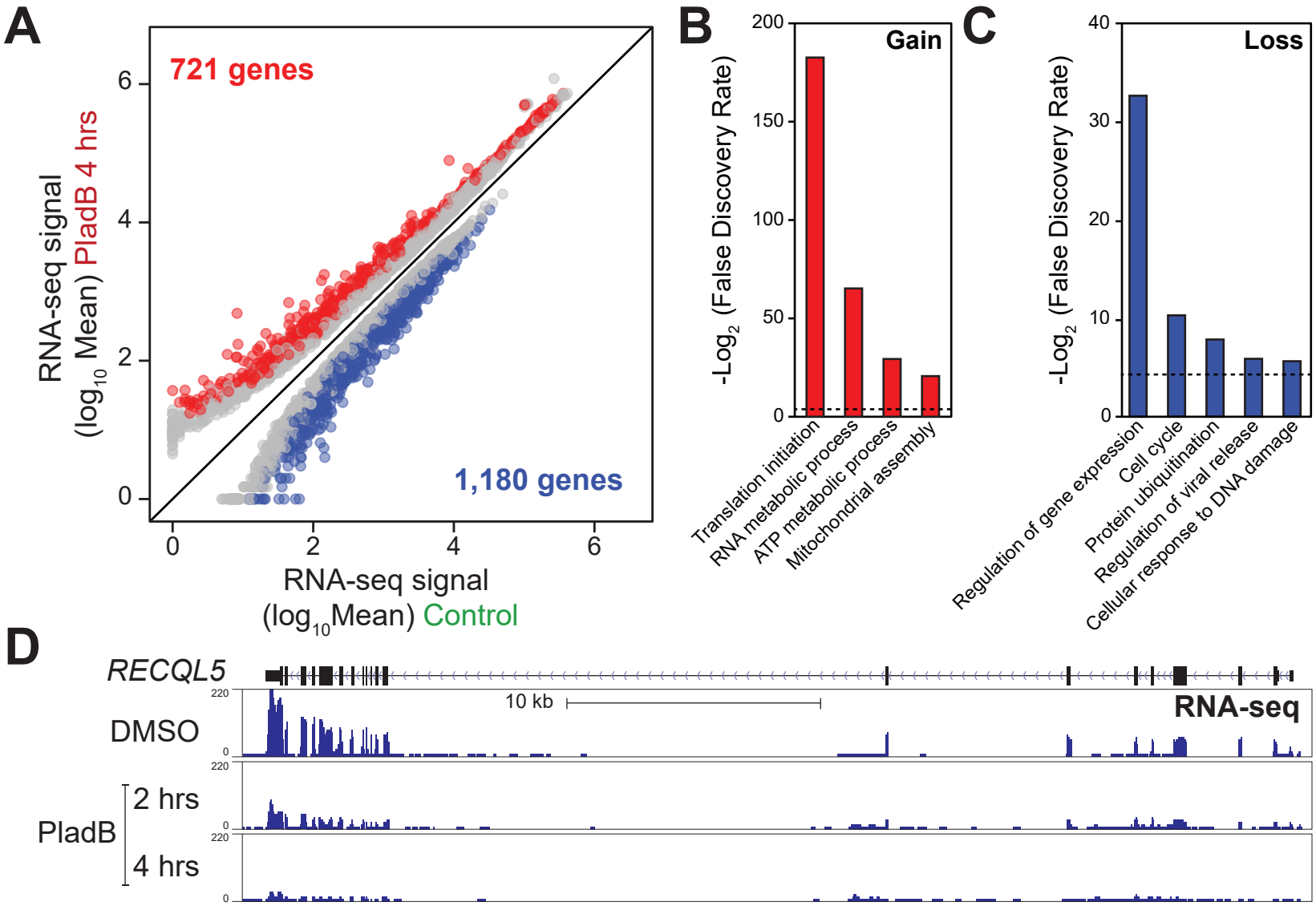
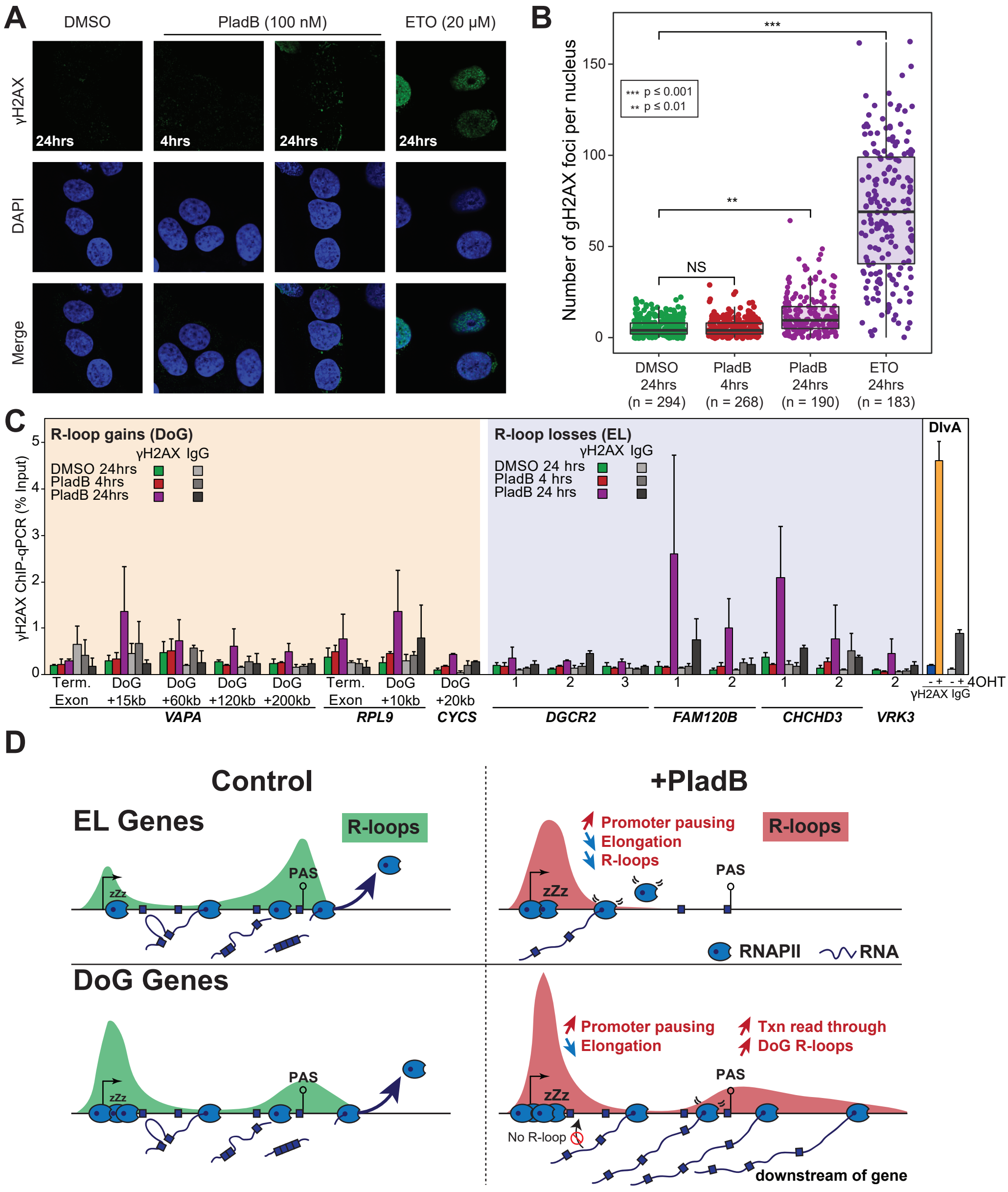

A
721 genes
1,180 genes
RNA-seq signal
(log10 Mean) PladB 4 hrs
RNA-seq signal
(log10Mean) Control
0
2
4
6
B
Gain
-Log2 (False Discovery Rate)
200
150
100
50
0
Translation initiation
RNA metabolic process
ATP metabolic process
Mitochondrial assembly
C
Loss
-Log2 (False Discovery Rate)
40
30
20
10
0
Regulation of gene expression
Cell cycle
Protein ubiquitination
Regulation of viral release
Cellular response to DNA damage
D
RECQL5
10 kb
RNA-seq
DMSO
PladB
2 hrs
4 hrs
220
0

Figure 7


A
DMSO
PladB (100 nM)
ETO (20 μM)
γH2AX
DAPI
Merge
24hrs
4hrs
24hrs
24hrs
B
Number of gH2AX foci per nucleus
*** p ≤ 0.001
** p ≤ 0.01
***
**
NS
DMSO 24hrs (n = 294)
PladB 4hrs (n = 268)
PladB 24hrs (n = 190)
ETO 24hrs (n = 183)
C
R-loop gains (DoG)
R-loop losses (EL)
DIvA
γH2AX ChIP-qPCR (% Input)
γH2AX IgG
DMSO 24hrs
PladB 4hrs
PladB 24hrs
DMSO 24 hrs
PladB 4 hrs
PladB 24 hrs
Term. DoG DoG DoG DoG Term. DoG DoG
Exon +15kb +60kb +120kb +200kb Exon +10kb +20kb
VAPA RPL9 CYCS
1 2 3 1 2 1 2 2
DGCR2 FAM120B CHCHD3 VRK3
- + - + 4OHT
γH2AX IgG
D
Control
+PladB
EL Genes
R-loops
PAS
zZz
Promoter pausing
Elongation
R-loops
RNAPII
RNA
DoG Genes
Promoter pausing
Elongation
Txn read through
DoG R-loops
No R-loop
downstream of gene