CDK12 loss in cancer cells affects DNA damage response genes through premature cleavage and polyadenylation

Malgorzata Krajewska[1,2,11], Ruben Dries[1,2,3,11], Andrew V. Grassetti[4], Sofia Dust[5], Yang Gao[1,2], Hao Huang[1,2], Bandana Sharma[1], Daniel S. Day[6], Nicholas Kwiatkowski[7], Monica Pomaville[1], Oliver Dodd[1], Edmond Chipumuro[1], Tinghu Zhang[7], Arno L. Greenleaf[8], Guo-Cheng Yuan[3,9], Nathanael S. Gray[7,10], Richard A. Young[6], Matthias Geyer[5], Scott A. Gerber[4], and Rani E. George[1,2,*]

[1]Department of Pediatric Hematology/Oncology, Dana-Farber Cancer Institute and Boston Children's Hospital, Boston, MA 02115, USA

[2]Department of Pediatrics, Harvard Medical School, Boston, MA 02115, USA

[3]Departments of Biostatistics and Computational Biology, Dana-Farber Cancer Institute, Boston, MA 02215, USA

[4]Department of Molecular and Systems Biology, Geisel School of Medicine at Dartmouth, Lebanon, NH 03756, USA

[5]Institute of Structural Biology, University of Bonn, 53127 Bonn, Germany

[6]Whitehead Institute for Biomedical Research, Massachusetts Institute of Technology, Cambridge, MA 02142,

[7]Department of Cancer Biology, Dana-Farber Cancer Institute, Boston, MA 02215, USA

[8]Department of Biochemistry, Duke University Medical Center, Durham, NC 27710, USA

[9]Harvard School of Public Health, Boston, MA 02115, USA

[10]Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston, MA 02115, USA

[11]these authors contributed equally

*Correspondence: Rani_George@dfci.harvard.edu

The cyclin-dependent kinase 12 (CDK12) modulates transcription elongation by phosphorylating the carboxy-terminal domain of RNA polymerase II and appears to selectively affect the expression of DNA damage response (DDR) and mRNA processing genes. Yet, the mechanism(s) by which it achieves this selectivity remains unclear. Using a highly selective CDK12/13 inhibitor, THZ531, and nascent RNA sequencing, we show that CDK12 inhibition results in gene length-dependent elongation defects, leading to premature cleavage and polyadenylation (PCPA) as well as loss of expression of long (>45 kb) genes, a substantial proportion of which participate in the DDR. This early termination phenotype correlated with an increased proportion of intronic polyadenylation sites, a feature that was especially prominent among DDR genes. Finally, phosphoproteomic analysis indicated that pre-mRNA processing factors, including those involved in PCPA, are direct phosphotargets of CDKs 12 and 13. These results support a model in which DDR genes are uniquely susceptible to CDK12 inhibition due primarily to their relatively longer lengths and lower ratios of U1 snRNP binding to intronic polyadenylation sites.

Eukaryotic gene transcription is facilitated by the orchestrated action of transcriptional cyclin-dependent kinases (CDKs) and associated pre-mRNA processing factors[1,2]. Transcriptional CDKs phosphorylate the carboxy-terminal domain (CTD) of RNA Polymerase II (Pol II) which serves as a platform for the recruitment of factors controlling transcriptional and post-transcriptional events. During transcription initiation, CDK7 (Kin28 in *S. cerevisiae)*, a subunit of TFIIH, phosphorylates serine 5 of the CTD[3]; subsequently, the release of paused Pol II and the transition to elongation is mediated by CDK9, a subunit of pTEFb, which phosphorylates the CTD at serine 2[4]. Studies in yeast and metazoans have shown that another transcriptional kinase, CDK12, together with its associating partner, cyclin K, modifies serine 2 of the Pol II CTD[5-7]. A second, less-studied metazoan ortholog of yeast Ctk1 in human cells is CDK13, which shares a largely conserved kinase domain with CDK12[6]. Although the biological role of CDK13 is not known, its sequence similarity with CDK12 predicts some degree of overlap between these kinases. In contrast to other transcriptional CDKs, both CDK12 and CDK13 contain additional arginine/serine-rich (RS) domains that are critical for proteins involved in processing premature RNA[8,9]. However, based on genetic depletion studies, CDK12 but not CDK13 has been reported to control the expression of DNA damage response (DDR) genes[6,10]. The selective regulation of these genes by CDK12 is also evident in cancers with loss-of-function *CDK12* mutations, such as high-grade serous ovarian carcinoma and metastatic castration-resistant prostate cancer, where a 'BRCAness' phenotype with genomic instability sensitizes cells to DNA cross-linking agents and poly (ADP-ribose) polymerase (PARP) inhibitors[11-14]. Similarly, suppression of wild-type CDK12 in Ewing sarcoma cells driven by the EWS/FLI fusion oncoprotein using THZ531[15] (a selective inhibitor of CDK12/13) also led to the decreased expression of DDR genes[16]. Hence, CDK12 loss of function, whether spontaneous or induced, appears to preferentially affect genes that have prominent roles in DNA repair and consequently in maintaining the stability of the genome.

Despite growing knowledge of CDK12 function in cancer cells and the availability of selective CDK12/13 inhibitors, the molecular basis for the selective effects of this kinase on DDR genes remains unclear. This deficit could have important implications for understanding distinctions among transcriptional CDKs and devising treatments for cancers that rely on aberrant transcription and/or genomic instability for their sustained survival and growth. Thus, using *MYCN*-amplified neuroblastoma (NB) as a solid tumor model characterized by

constitutive transcriptional upregulation[17], and genomic instability[18,19], but lacking CDK12 mutation[20], we demonstrate a mechanistic link between the structural properties of DDR genes and their susceptibility to the suppressive effects of CDK12 inhibition.

Results

CDK12/13 inhibition with THZ531 is selectively cytotoxic in NB cells and affects distal transcriptional elongation. To understand the preferential effect of CDK12 on the DDR, we first determined whether we could abrogate its activity by using THZ531. This covalent inhibitor inactivates both CDK12 and 13, albeit at different potencies (IC_{50} against CDK12, 158 nM; CDK13, 69 nM)[15], and is 50-fold more active against CDK12/13 than CDK7 and CDK9, both of which have been therapeutically targeted in cancer[17,21]. THZ531 binds to unique cysteine residues outside the canonical kinase domains of both CDK12 and 13 (Cys1039 and Cys1017, respectively), resulting in their prolonged and irreversible inactivation[15]. It does not covalently target CDK9 due to the absence of a unique cysteine in this position.

We observed strong selectivity and cytotoxicity in NB cells compared to nontransformed cells (**Fig. 1a, Supplementary Fig. 1a**). Decreased sensitivity was also observed in Kelly E9R NB cells expressing a point mutation at the CDK12 Cys1039 THZ531 binding site[22] [IC_{50} = 400 nM compared to 60 nM in Kelly wild-type (WT) cells], suggesting that inhibition of CDK13 alone did not affect cell viability. Moreover, target engagement studies using a biotinylated derivative of the compound (bio-THZ531) revealed consistently decreased binding to CDK12 and CDK13 after treatment with THZ531, indicating that these kinases are indeed targets of this inhibitor **(Supplementary Fig. 1b).** THZ531 treatment led to apoptosis as well as G2/M cell cycle arrest in these cells (**Supplementary Fig. 1c-e)**. The sensitivity to THZ531 extended to both *MYCN*-amplified and nonamplified NB cells; in the latter, the addition of an ABCB1 drug efflux pump inhibitor (tariquidar) was necessary to overcome high expression of this protein and subsequent inhibitor efflux[22,23] (**Supplementary Fig. 1f)**. Despite the role of CDK12 in transcription elongation[24], THZ531 induced variable dose- and time-dependent decreases in Pol II Ser2 phosphorylation, with minimal effects on Ser5 or 7 phosphorylation (**Fig. 1b**). However, we observed striking downregulation of termination–associated Pol II threonine (Thr4)[25-27] phosphorylation, indicating that distal elongation was affected (**Fig. 1b**). Together, these results indicate that THZ531, by binding to CDK12/13, induces cytotoxicity in NB cells through effects on transcription elongation.

CDK12 inhibition preferentially affects DDR genes. CDK12 inhibition has been shown to affect the expression of genes involved in the DDR[5,6,10]. To determine whether similar effects are produced by our selective inhibitor in NB cells, we analyzed the gene expression profiles of cells treated with and without THZ531 for 6 hr., a time point at which there were little or no confounding effects due to cell cycle changes (**Supplementary Fig. 1e**). Unlike the effects seen with THZ1[17], predominantly an inhibitor of CDK7 with some activity against CDK12/13[28], we failed to observe a complete and global transcriptional shutdown in THZ531-treated NB cells; instead, only 57.4% of the transcripts were downregulated (n=10,707), with 0.35% (n=66) upregulated [false discovery rate (FDR) <0.05] (**Supplementary Fig. 2a; Supplementary Table 1**). Consistent with earlier studies[15,16], THZ531 led to significant downregulation of both transcription-associated and DDR genes (**Fig. 1c, Supplementary Fig. 2b, c**), the latter of which were primarily associated with homologous recombination (HR) repair and are crucial for the maintenance of genome stability, including *BRCA1, BARD1* and *RAD51*[29] (**Fig. 1c, Supplementary Fig. 2c, d)**. To determine whether these effects were due to inhibition of CDK12 or 13, we depleted the expression of each kinase individually in NB cells, and in keeping with prior studies[6,10,30], observed selective downregulation of DDR genes with CDK12 and not CDK13 knockdown (KD) (**Supplementary Fig. 2e**). Additionally, the expression of DDR genes was not affected in Kelly E9R THZ531-resistant cells, further implicating the selective role of CDK12 in regulating the DDR (**Fig. 1d**). Consistent with these observations, THZ531 also led to increased DNA damage with elevated γ-H2AX levels (**Fig. 1e)** and decreased radiation-induced RAD51 foci, indicating defects in DNA repair **(Fig. 1f)**. Thus, our findings indicate that DDR genes are selectively affected by THZ531 and that such regulation is driven predominantly by CDK12.

CDK12/13 inhibition with THZ531 leads to an elongation defect. The transcriptional effects of CDK12/13 inhibition with THZ531, including downregulation of the steady-state expression of DDR genes, occurred as early as 6 hrs. post-treatment and independently of cell cycle changes (**Supplementary Fig. 1e**). This result, plus the fact that CDK12 and 13 have been implicated in pre-mRNA processing[10,31,32] where observable changes are likely to occur within minutes or hours, indicated that further analysis of steady-state RNA would not be sufficient to fully characterize alterations in the tightly coupled Pol II transcription and RNA processing processes or

discriminate between early and late effects due to CDK12 perturbation. Hence, we used transient transcriptome sequencing (TT-seq), a modification of the 4-thiouridine (4sU)-pulse labeling method[33], with spike-in controls for normalization of input RNA amount to identify the immediate changes in nascent RNA production in cells exposed to THZ531 for 30 min and 2 h. The data show adequate exonic and intronic coverage, including that of 5′ upstream and 3′ downstream regions outside annotated transcripts (mean 59% of reads to introns, 35% to exons and 6% to flanking regions), indicating that nascent RNAs, including a large proportion of preprocessed RNAs were captured (**Supplementary Fig. 3a**). Altogether, we detected 12,260 protein coding, 4,809 long non-coding and 3,816 short non-coding genes (transcripts per million > 2). At 30 min post-treatment, several immediate-early response gene transcripts were induced, thus confirming the ability of TT-seq to detect early changes in transcription, but this effect was not sustained at 2 h (**Supplementary Fig. 3b; Supplementary Table 2**). Instead, the changes in nascent transcription were more pronounced, leading us to focus our analyses on this time point. DDR genes comprised one of the downregulated gene groups (**Supplementary Table 3; Supplementary Fig. 3c, d**), consistent with our gene expression profiling of steady-state RNA, which demonstrated downregulation of these genes after a 6-h treatment with THZ531 (**Fig. 1c** and **Supplementary Fig. 2a-c**). We validated this result by measuring nascent RNA expression along the *BRCA1* gene by qRT-PCR, observing a gradual decline in expression from the 5′ to the 3′ end of the gene following THZ531 treatment (**Supplementary Fig. 3e**). Gene ontology (GO) enrichment analysis of the top 400 most downregulated genes also revealed genes associated with transcription and mRNA processing (**Supplementary Fig. 3f**).

To further elucidate the effect of CDK12/13 inhibition on RNA synthesis, we first analyzed the changes in nascent RNA expression over gene bodies. Average meta-gene analysis of protein-coding genes and all classes of long noncoding RNAs demonstrated a prominence of TT-seq signals both upstream and downstream of the transcription start sites (TSS) (**Fig. 2a**). Since increased Pol II pausing has recently been shown to inhibit new initiation[34,35], we next asked whether pausing was affected by CDK12/13 inhibition by calculating the change in nascent transcript read density over regions flanking the TSS (-500 bp to 1000 bp) following THZ531 treatment. This analysis showed a gradual increase in TT-seq reads, with peak signal accumulation occurring 1000 bp

downstream of the TSS -- well beyond known Pol II pausing sites (30-100 bp) **(Fig. 2b, Supplementary Fig. 4a)**. Moreover, the rate at which the change in TT-seq signals occurred following THZ531 treatment continued to increase up to 250 bp beyond the TSS, after which a decrease was seen **(Fig. 2b)**. Together, these observations suggest that THZ531 treatment does not lead to defective Pol II pause release; in fact, in keeping with the recently proposed model[34,35], pause release may even be increased, which in turn would account for the observed increase in initiation. The finding that upstream antisense RNAs (uaRNAs) (which are short-lived and do not undergo extensive processing) were also increased at the TSS **(Fig. 2a)** supports this notion. After the initial 5′ increase in read density, a rapid loss of reads from the 5′- to the 3′-ends of genes was seen (**Fig. 2a**), with a net average loss of read density of around 10 kb 3′ of the TSS **(Supplementary Fig. 4a)**. Together, these findings point to an elongation defect upon CDK12/13 inhibition.

The THZ531-induced elongation defect is dependent on gene length. Because of the wide range in gene lengths throughout the genome (<1kb to >1Mb) and prior reports that CDK12 preferentially regulates the expression of long genes[36], we next determined whether this variable had any effect on the elongation defect seen with the CDK12/13 inhibitor. Notably, there was a significant correlation between gene length and downregulation of gene expression: the longer the gene, the more likely it was to be downregulated **(Fig. 2c)**. To define this relationship further, we divided the downregulated genes into 4 quartiles based on the distribution of gene lengths [short (< 9.9 kb), medium-short (9.9 - 26.4 kb), medium-long (26.4 - 64.5 kb) and long (> 64.5 kb)]. As shown in **Fig. 2d**, the long genes consistently had the most pronounced elongation defect and, concomitantly, the greatest transcriptional downregulation. When we restricted our differential gene expression analysis to protein-coding genes and TT-seq reads that fell within exonic regions, and compared the results against these unbiased gene length groups, we observed that 362 (7%) of 5110 longer genes (202 long and 160 medium-long) were downregulated, while only 111 (2%) of 4895 shorter genes (14 medium-short and 97 short) were upregulated (adjusted $p < 0.05$; log2 fold change < -1). GO analysis of these genes showed that the top categories comprised DDR genes **(Fig. 2e)**. This length-dependent elongation defect was not observed either for long or short noncoding RNAs, although global downregulation of the latter was observed (**Supplementary Fig. 4b**). In contrast to longer

genes, we observed that short or very-short (<3.4 kb) genes were upregulated or transcribed normally and showed significant 3′ UTR increase and extension (**Supplementary Fig. 4c)**. This subgroup was enriched for replication-dependent (RD) histone genes (n=72) (**Fig. 2f)** which do not normally rely on polyadenylation for transcription termination but on stem-loop binding[37,38], a process regulated by Pol II Thr4P that was profoundly decreased following THZ531 treatment (**Supplementary Fig. 4d, Fig. 1b**). This effect (validated by qRT-PCR of total and nascent RNA) was also prominent among the amplified and overexpressed *MYCN* oncogene (6.4 kb long) and explained its lack of downregulation due to THZ531 in *MYCN*-amplified cells (**Supplementary Fig 4e, f)**. Together, these results suggest that CDK12/13 inhibition with THZ531 leads to an elongation defect which predominantly involves genes within the longer length category, with normal or increased expression of shorter genes.

CDK12 inhibition leads to premature cleavage and polyadenylation. The gradual decrease in nascent RNA expression from the 5′ to the 3′ ends of long genes with THZ531 treatment implied a possible termination defect. To pursue this notion, we performed global poly(A) 3′-sequencing of cells treated with THZ531. The majority (69%) of the identified poly(A) peaks were associated with known upstream polyadenylation site (PAS) motifs, the most abundant being the canonical AATAAA motif (**Supplementary Fig. 5a**). In agreement with their nascent RNA expression profiles **(Fig. 2e, f)**, transcripts associated with long genes showed a loss of annotated terminal or 3′ poly(A) sites **(Fig. 3a,** left**)**, while short transcripts, such as those of the histone processing genes and *MYCN,* terminated at distal unannotated poly(A) sites (**Fig. 3a,** right**, Supplementary Fig. 4e).** These findings suggest a transcription termination defect following CDK12/13 inhibition, with differential effects based on gene length.

Given the decrease in annotated terminal poly(A) sites, we attributed the genome-wide elongation and termination defects in THZ531-treated cells to premature cleavage and polyadenylation (PCPA). Transcription of most protein-coding genes is terminated at the 3′ ends of genes through cleavage and polyadenylation, whereas premature transcription termination such as PCPA occurs within a short distance from the TSS and results in the production of aberrant transcripts[39-41]. We therefore analyzed the distribution of poly(A) 3′-seq reads for all

protein-coding genes across the genome, observing a time-dependent increase in polyadenylated sites at the 5′ proximal ends of genes following CDK12 inhibition (**Fig. 3b**). Further study of the poly(A) sites that were differentially utilized between THZ531- and DMSO-treated cells focused on those that showed at least a 2-fold change (**Supplementary Fig. 5b**). This approach revealed a strikingly different 3'-peak distribution in THZ531-treated compared with DMSO-treated cells, with significant enrichment in intronic regions (60% vs. 36%) and an almost complete absence of reads at annotated transcription end sites (TES) (4% in THZ531- vs. 20% in DMSO-treated cells; **Supplementary Fig. 5c**), an effect that was most prominent for the top 5000 differential poly(A) peaks between the two samples **(Fig 3c)**. Together, these findings indicate that CDK12/13 inhibition in NB cells causes generalized PCPA with the use of cryptic intronic polyadenylation sites. Interestingly, the THZ531-induced effect at the nascent RNA level was computationally inferred[42] as early as 2 h post treatment, with PCPA apparent in 809 (7%) of the 11,902 protein-coding genes containing at least one intron (**Fig 3d**). Poly(A) 3′-seq data showed that more than half of these genes underwent early termination in the first two introns/exons (59%, 476/809) and almost three-quarters in the first four introns/exons (73%, 587/809) (**Fig 3d**). Integrative analysis of TT-seq and poly(A) 3′-seq data at the 5′ proximal regions (-1 kb to +1 0kb of TSS) showed that the aberrant accumulation of 5′ proximal TT-seq reads coincide with the peaks of proximal 3′ poly(A) usage, implying that most transcripts are terminated early at the beginning of elongation (**Supplementary Fig. 5d**). This inference was further supported by the correlation between decreased nascent reads along the 5′ to 3′regions and the usage of proximal poly(A) sites in THZ531-treated cells, suggesting a high probability of proximal poly(A) site usage that gradually diminishes when elongation is terminated due to PCPA.

The termination defect seen with THZ531 is due to inhibition of CDK12. We next asked whether the observed effects on termination through PCPA could be assigned specifically to CDK12 or 13 by genetic depletion (shRNA KD) followed by poly(A) 3′-sequencing. CDK12-depleted cells displayed the highest and most significant increase in poly(A) 3′-sequencing reads at the 5′ proximal ends of genes compared to control shRNA-expressing cells (**Supplementary Fig 5e**). Although depletion of CDK13 also resulted in an increase in 5′ proximal reads, this effect was significantly lower than that seen with CDK12 depletion (**Supplementary Fig 5e**). Only CDK12-

depleted cells showed an increased usage of intronic poly(A) sites; this phenomenon was not evident in CDK13-depleted cells (**Supplementary Fig 5f)**. Importantly, THZ531 treatment in Kelly E9R cells with the THZ531-binding site mutation did not display any increase in 5′ proximal reads (**Fig. 3e**) or in intronic poly(A) site usage compared to wild-type Kelly cells (**Fig. 3f**), suggesting that targeting of CDK13 alone was not sufficient to induce the PCPA defect. We had previously observed that THZ531 treatment led to a gene length-dependent decrease in nascent RNA expression (**Fig. 2c).** This length-dependent effect translated into decreased gene expression as well in cells with CDK12 shRNA depletion, but not in cells with CDK13 shRNA KD or in E9R cells treated with THZ531 (**Supplementary Fig. 5g**). Together, these results further identify PCPA as the main defect resulting from THZ531 treatment, an outcome that is mediated primarily by its targeting of CDK12.

CDK12/13 inhibition induces minimal splicing alterations in NB cells. Because previous studies point to a role for CDK12 in splicing regulation[10,32,43,44], we determined whether aberrant splicing could explain the elongation defect seen with THZ531 treatment. Analysis of the nascent transcriptomic data showed that in general, there was a paucity of significantly altered splicing events following THZ531 treatment. The largest proportion of splicing defects comprised intron retention (13.4%), followed by alternative 5′ and 3′ splicing (4.7% and 4.8% respectively), while skipped and mutually exclusive exons were rarely observed (**Fig. 4a**). To further investigate intron retention, we calculated the intron retention (IR) index (log2 ratio of intron vs. exon TT-seq signal coverage differences between THZ531- and DMSO-treated cells; see Methods), and noted overall intron loss (642 of 11,155 protein-coding genes, 5.7%) together with a low exon/intron length ratio (IR<1) **(Fig. 4b)** in genes that were downregulated by THZ531. Importantly, this effect was seen primarily at long genes. Short genes, on the other hand, were characterized by intron retention (156 of 11,155 genes, 1.3%) and a high exon/intron length ratio (IR>1, adjusted $p < 0.05$) (**Fig. 4c**). We reasoned that the apparent increased splicing efficiency in long genes was likely not due to a more efficient spliceosome, but rather, was a secondary effect of the severe elongation defect seen within these genes (**Fig. 2a, d**). To pursue this hypothesis, we calculated the individual IR indices for the combination of the first exon/intron and last exon/intron length-ratios of the long genes that displayed intron loss, observing a greater intron loss for the last exon/intron compared to that of the first

exon/intron (**Fig. 4d**). These results suggest that the lack of intron coverage at the 3′ end in longer genes was likely due to defective elongation together with the reduced formation of such long transcripts following THZ531 treatment.

Gene length and a lower U1 snRNP/PAS ratio predispose DDR genes to PCPA. Genes that underwent THZ531-induced PCPA were significantly longer than genes that did not undergo this change, as might be expected from the elongation defect in the long gene group (>64.5 kb; **Fig. 2d, Supplementary Fig. 6a**). Importantly, the group of long genes that underwent PCPA was specifically enriched for DDR genes, such as *BARD1*and *BLM*, with respective lengths of 84 kb and 98 kb; **Fig 5a-c, Supplementary Fig. 6b).** We validated this finding through 3′ RACE of the *BARD1* transcript in THZ531-treated cells **(Supplementary Fig. 6c)**. Interestingly, we noted that DDR genes undergoing PCPA as a result of CDK12 inhibition had a statistically higher number of intronic poly(A) sites relative to other genes of similar length (**Fig. 5d),** indicating that gene length alone does not fully explain the specific vulnerability of this subset of genes to early termination. Hence, to assess the relative contribution of gene length to the early termination phenotype observed after THZ531 treatment, we tested other determinants known to influence co-transcriptional processing[41,45,46]. Apart from longer gene length, we noted that a longer first intron, a larger number of introns, higher gene expression, lower GC content and a lower U1 snRNP/PAS ratio were also associated with early termination due to PCPA, with the latter two features emerging as the most significant based on effect size (**Fig 5d, e**). The combination of these two features is not surprising given that the U1 snRNP/PAS ratio is largely determined by DNA sequence, and that these variables showed the highest pairwise correlation (data not shown).

The U1 snRNP complex prevents premature termination through recognition and inhibition of cryptic poly(A) sites[40-42,47]. Indeed, Oh *et al*[42] demonstrated that direct depletion of U1 in HeLa cells using morpholino KD results in the decreased expression of long genes. We observed a significant overlap between genes that underwent PCPA in this data set and those that were similarly affected by THZ531 treatment, even though they represent two different cancer cell types and were studied at different time points after perturbation of different

targets – U1 at 4 and 8 h[42] and CDK12 at 2 h (this study) **(Supplementary Fig. 7a; Supplementary Table 4)**. This finding is supported by the significantly increased usage of intronic poly(A) sites in DDR genes, even when compared with the genome-wide increase that was observed following THZ531 treatment in wild-type Kelly NB cells (**Supplementary Fig. 7b,** left**; Fig. 3f**). Importantly, no such change was seen in Kelly E9R THZ531-resistant cells (**Supplementary Fig. 7b,** right**)**. In addition, genes that showed increased intronic poly(A) site usage following THZ531 exposure were enriched for GO categories associated with DNA damage **(Supplementary Fig. 7c)**. Finally, the expression of DDR genes following THZ531 treatment was significantly reduced in WT cells compared to E9R cells expressing the Cys1039 mutation, further supporting the role of CDK12 in regulating the expression of DDR genes (**Supplementary Fig. 7d, e**). In conclusion, these observations indicate that CDK12 inhibition leads to premature termination that depends on gene length and the U1 snRNP/PAS ratio and may provide an explanation for the selective effects of this transcriptional kinase on DDR gene expression (**Fig. 5f**).

CDK12/13 phosphorylates RNA processing proteins. Our results demonstrate the effect of CDK12 inhibition on transcription elongation and identify PCPA as a potential explanation for this selectivity. Given that the transcriptional activity of Pol II and processing of nascent transcripts occur simultaneously[2], we hypothesized that the CDK12 and/or 13 kinases may regulate the phosphorylation of targets other than the Pol II CTD, which could contribute to cotranscriptional RNA processing. To address this question, we performed phosphoproteomics analyses of cells treated with and without THZ531 using stable isotope labeling with amino acids in cell culture (SILAC). This study revealed a ≥2-fold increase of 88 phosphopeptides and a similar decrease in 129 sites ($p < 0.1$; Student's *t*-test; **Fig. 6a, Supplementary Table 5**). The majority of phosphorylation sites that decreased in abundance upon THZ531 treatment occurred at serine or threonine residues, usually with a proline in the +1 position - the minimal consensus recognition site for all CDKs[48] **(Supplementary Fig. 8a)**. Protein interaction network analysis of all identified substrates clustered into two groups, the larger of which contained phosphorylated proteins centered on Pol II, while the other consisted of phosphorylated proteins that interact directly with CDK12 **(Supplementary Fig. 8b)**. Interestingly, proteins encoded by DDR genes were not

significantly represented in this analysis, suggesting that CDK12 may not directly regulate DDR protein phosphorylation.

GO analysis of candidate CDK12 substrates that were significantly decreased in abundance after THZ531 treatment revealed mRNA processing factors as the top category, accounting for more than 50% of the identified phosphoproteins **(Fig. 6b).** Interestingly, one of the top mRNA processing factors was the small nuclear ribonucleoprotein SNRNP70, which associates with U1 as part of the U1 snRNP complex[49] (**Fig. 6a).** Other top phosphoproteins that were affected by CDK12/13 inhibition included the PRP19 complex protein[50,51], CDC5L with roles in RNA splicing and genomic stability; SF3B1, a component of the splicing machinery that is involved in pre-mRNA splicing[52]; and SPT6H, which controls transcription elongation rate and release of paused Pol II into productive elongation[53]. We confirmed the phosphorylation of these candidates using ^{32}P-labeled ATP *in vitro* kinase assays using GST-tagged substrates together with CDK12/CycK and CDK13/CycK **(Supplementary Fig. 8c).** Similar to CDK12, CDK13 phosphorylated the substrate proteins in a time-dependent manner **(Fig. 6c, Supplementary Fig. 8d).** Of note, CDK12-mediated phosphorylation resulted in a higher rate of ^{32}P incorporation for CDC5L and SF3B1, suggesting that CDK12 phosphorylates more sites in these substrates than CDK13 (**Fig. 6c**). Additionally, control experiments without the addition of either kinase revealed that phosphorylation of CDC5L and SF3B1 was significantly below that measured in presence of the active kinases. The target measurement with SPT6H, however, showed no increase in radioactive counts by either kinase compared to the control measurement without kinase, most likely because the truncated recombinant protein (GST-SPT6H 1323-1544) was used in the kinase assay, rather than the full-length protein in its native cellular protein-complex (**Supplementary Fig. 8d)**. Next, we repeated the kinase assays after pre-treatment of the CDK/cyclin complex with THZ531, noting reduced phosphorylation of the CDC5L and SF3B1 substrate proteins with increasing concentrations of the inhibitor (**Fig. 6d**). Finally, to identify the exact sites phosphorylated by CDK12/CycK in the in vitro kinase assays, we performed peptide mass fingerprint analyses of the recombinant protein substrates, which confirmed the following phosphorylation sites identified in the SILAC analysis: CDC5L (pT396), SF3B1 (pT326), as well as SPT6H (pT1523, **Supplementary Fig. 8e**). Together, these results suggest that these candidate proteins are substrates of CDK12/13.

The phosphorylation of RNA processing factors by CDK12 is consistent with their arginine/serine-rich (RS) domains, typical of RNA-interacting and splicing factors[8] which also reside within nuclear speckles[54]. Indeed, we observed enlarged nuclear speckles ('mega speckles') in cells treated with THZ531, similar to those formed after treatment with the splicing inhibitor pladienolide B **(Fig. 6e, Supplementary Fig. 8f)**, suggesting that the underutilization of proteins involved in mRNA processing upon CDK12 inhibition causes their retention in these bodies. Thus, both CDK12 and 13 phosphorylate pre-mRNA processing factors that could affect their recruitment to Pol II.

DISCUSSION

In this study, we took advantage of the selectivity and irreversibility of a covalent inhibitor of CDKs 12 and 13 to dissect the dynamic and immediate alterations in co-transcriptional RNA processing in NB cells. Using nascent RNA and poly(A) 3′-sequencing, we demonstrate that such inhibition leads to a gene length-dependent elongation defect associated with early termination through PCPA **(Fig. 6f).** Especially vulnerable to this defect were long genes with a lower ratio of U1 snRNP binding to poly(A) sites, which include many of those involved in the DDR. Conversely, short genes showed an increased likelihood of intron retention and 3′ UTR extension, or, as in the case of the non-polyadenylated replication-dependent histone genes, the generation of polyadenylated transcripts.

We further identified CDK12 as the predominant kinase mediating the observed transcriptional effects of THZ531 in treated cells. However, both CDK12 and CDK13 appear to induce the phosphorylation of RNA processing proteins that have established roles in cotranscriptional RNA processing by Pol II. THZ531 inhibits both CDK12 and CDK13[15], because the cysteine (Cys1039) outside the CDK12 kinase domain that is being targeted for covalent linkage is also present in CDK13 (Cys1017); importantly, no other transcriptional CDK contains a cysteine in a similar position[15]. We also observed that E9R[10] NB cells harboring a point mutation at the CDK12 Cys1039 THZ531 binding site were less sensitive to THZ531 than were WT cells. Moreover, THZ531 treatment of E9R cells revealed significantly fewer length-dependent elongation defects and PCPA, compared to findings in cells expressing WT CDK12. The distinction between phosphorylation targets, however, was not as clear-cut; both CDK12 and CDK13 induced the phosphorylation of RNA processing proteins, although further studies are required to fully resolve this issue.

Inhibition of CDK12 by THZ531 results in increased transcription initiation, supporting a model in which Pol II undergoes cycles of abortive release after initiation due to premature termination in the elongation phase[55]. This conclusion is supported by our data showing the accumulation of TT-seq reads both upstream antisense, and those at the TSS that go well beyond the point of Pol II pausing (average peak observed at ~1000 nt downstream of the

TSS) and by the increased *rate* of signal accumulation at the Pol II pause sites themselves. We also observed a sharp decrease in the accumulation of TT-seq reads at the start of elongation, suggesting that the Pol II complex is released from chromatin with resultant early termination during the elongation phase. This observation is further supported by the increase in poly(A) 3'-seq reads at the 5′ proximal ends of these genes. Thus, we postulate that in the absence of CDK12 activity, Pol II is unable to enter into productive elongation; instead, as previously proposed[55], it is released from chromatin, likely increasing the pool of free Pol II molecules that can engage in transcription initiation, which, in turn leads to the increased TT-seq reads observed at the TSS in THZ531-treated cells. Thus, transcription initiation is increased, followed by decreased Pol II pausing and/or increased Pol II pause release, but with a failure to transition to productive elongation leading to early termination due to PCPA.

These findings agree with - but differ mechanistically from – those reported with inhibition of the other Pol II Ser2 elongation kinase, CDK9, where increased Pol II pausing leads to a defect in elongation and negatively impacts transcription initiation[34,35]. By contrast, perturbation of CDK12 results not only in a defect in elongation, but also, *increased* transcription initiation and possibly Pol II pause release. Thus, CDK12 and CDK9 inhibition have opposite effects on both transcription initiation and Pol II pausing, underscoring again that the observed effects in this study are mediated predominantly through the selective inhibition of CDK12, and that CDK9 and CDK12 have distinct roles in transcription regulation.

The RNA processing defects seen upon CDK12 inhibition, including the alternative use of poly(A) sites and polyadenylation of replication-dependent histone mRNAs, are likely due to slowing of the elongating RNA Pol II complex. Indeed, it was recently demonstrated that the regulation of transcription speed modulates pre-mRNA processing[56]; factors that decelerate Pol II elongation, such as UV irradiation, can generate long polyadenylated histone transcripts due to failure of stem-loop folding and subsequent failure to process 3′ ends. The profound decrease in phosphorylated RNA Pol II Thr4 in THZ531-treated cells (required for the recruitment of the stem-loop binding protein and other 3′ processing factors to replication-dependent histone genes)[57] supports this conclusion.

One plausible explanation of how CDK12 could accelerate Pol II elongation would be by directly facilitating mRNA processing steps during transcription. Indeed, this seems likely as the target proteins of CDK12-mediated phosphorylation are highly enriched for several of these factors and phosphorylation is known to provide a structural basis for the regulation of RNA-binding proteins[58] or for their targeted distribution within the nucleus[59]. We observed that in addition to gene length, a main determinant of premature termination was the U1 snRNP/PAS ratio, which was lower in DDR genes that underwent PCPA. It is well established that the U1 snRNP facilitates the transcription of long genes with its inhibition resulting in PCPA[42]. SNRNP70, a component of the U1 snRNP complex was identified as a potential phosphorylation substrate of CDK12/13 in our study; hence, it is quite possible that its decreased phosphorylation could partly account for the increased usage of alternate polyadenylation sites in DDR and other long genes. This could also explain why in contrast to findings in other studies implicating CDK12 in splicing regulation[32,43], CDK12 inhibition did not lead to major splicing alterations, most likely because transcription was terminated well before it reached the 3′ splice sites[53].

As shown schematically in **Fig. 6f,** we propose that CDK12 inhibition leads to a slowing of Pol II elongation and hence to an increase in the probability of using cryptic intronic poly(A) sites and undergoing PCPA. As such, long genes with low U1 snRNP/PAS ratios, such as DDR genes, are especially vulnerable to this loss, yielding an aborted elongation phenotype, manifested at the 3′ ends of these genes. Most importantly, our analysis demonstrates that CDK12 by itself lacks any intrinsic preference for DDR genes; instead, the properties of the gene target determine its sensitivity to CDK12 inhibition, and many DDR genes possess the requisite features. Not only was gene length a significant contributor to the PCPA phenotype, but our identified DDR genes harbored more intronic poly(A) sites than expected based on their longer gene lengths. DDR genes that evaded PCPA were those with genetic determinants that did not favor this process - such as shorter length, a short first intron and decreased numbers of introns. Future work is needed to resolve why so many genes involved in DNA repair have this genetic composition compared to the genome-wide background.

DNA damage in cancer cells has been shown to lead to widespread truncated transcripts and intronic PCPA at the 5′ ends of genes that are involved in the DDR. This response correlates with decreased U1 snRNP levels,

with the resultant intronic polyadenylation[60]. Conceivably, the effects we attribute to CDK12 inhibition may not be limited to cells with underlying DNA damage such as NB and possibly Ewing sarcoma[16], but also in cancers with CDK12 loss-of-function mutations such as ovarian cancer. Indeed, as recently demonstrated in a subset of prostate cancers with CDK12 loss-of-function mutations[14], the PCPA, as well as intron retention, observed with CDK12 inhibition could facilitate the formation of neoantigens that could be exploited to improve immune therapies or to develop personalized cancer vaccines[61]. Conversely, PCPA can also directly lead to the inactivation of tumor suppressor genes[62]; thus, further studies are necessary to understand and perhaps predict the frequency of PCPA within the genome.

In conclusion, the findings presented here advance understanding of the role of CDK12 in transcription regulation – of DDR genes in particular. Thus, by inducing an RNA Pol II elongation defect and subsequent usage of proximal poly(A) sites leading to premature cleavage and polyadenylation of long DDR genes, it was possible to decipher the mechanism by which THZ531 selectively abolishes the DDR in *MYCN*-driven NB cells which are highly dependent on adequate DNA repair function for their survival. The extent to which these observations apply to other genomically unstable cancers lacking CDK12 loss-of-function mutations is unclear; consequently, their verification in additional cancer models will be pivotal in generating molecular rationales for the therapeutic targeting of CDK12 across a broad cross-section of vulnerable tumors.

Acknowledgements We thank K. Adelman, T. Henriques, P. Cramer, S. Gressel and A. Meyer for detailed protocols, advice on experimental design and helpful discussions. We thank the Whitehead Institute Genome Core for RNA sequencing. We thank members of our laboratory for helpful discussions. This work was supported by NIH R01CA197336 (to R.E.G and R.A.Y), Friends for Life Neuroblastoma Foundation (to R.E.G), Alex's Lemonade Stand Foundation (to R.E.G and O.B.D.), the DFCI Claudia Adams Barr Award (to R.E.G) and the Rally Foundation for Childhood Cancer Research (to M.K.). The Gray laboratory is supported by the DFCI Hale Family Pancreatic Center and NIH grants, CA154303-06 and CA179483-02.

Author contributions M.K. designed and carried out molecular, cellular and genomic experiments. R.D. performed bioinformatics analyses. A.V.G. performed the SILAC experiments under the supervision of S.A.G. S.D. performed the in vitro kinase assays under the supervision of M.G. D.S.D. analyzed the microarray expression data. Y.G. established Kelly E9R cells and performed target engagement experiments with CDK13. N.K. provided protocols and assistance with target engagement studies. B.S. performed 3'RACE PCR and provided technical support. E.C. performed the initial cell viability assays of THZ531. H.H. generated the consensus sequences from the SILAC data. O. B. D. performed the combination studies under the supervision of M.K. T.H.Z. generated THZ531. A.L.G. provided CDK12 antibody and advice. G-C.Y. supervised the bioinformatics analyses. N.S.G. and R.A.Y. provided feedback on study design and experimental results. M.K., R.D. and R.E.G. wrote the manuscript. R.E.G. conceived the project and supervised the research. All authors discussed the results and commented on the manuscript.

Figure Legends

Fig. 1. CDK12/13 inhibition results in selective cytotoxicity in NB cells and affects distal transcription elongation. a, Dose-response curves for human NB cells treated with increasing concentrations of THZ531 for 72 h. Kelly E9R cells, which express a homozygous mutation at the Cys1039 THZ531-binding site in CDK12[22] (see Methods) were also included. Fibroblast cells (NIH-3T3, IMR-90, BJ) were used as controls. Cytotoxicity is reported as percent cell viability relative to DMSO-treated cells. Data represent mean ± SD; n=3. **b**, Western blot analysis of Pol II phosphorylation in NB cells treated with THZ531 or DMSO at the indicated concentrations for the indicated times. **c**, Waterfall plot of fold-change in gene expression in IMR-32 NB cells treated with THZ531, 400nM for 6 h; selected DDR genes are highlighted. **d,** qRT-PCR analysis of the indicated DDR gene expression in Kelly WT (left) cells and Kelly E9R (right), treated with THZ531 or DMSO at the indicated concentrations for 6 h. Data normalized to GAPDH are presented as mean ± SD; n=3**.** **e**, Flow cytometry analysis of γ-H2AX staining in Kelly NB cells treated with 400 nM THZ531 for the indicated time points (left). Quantification of staining (right) reported as mean ± SD of n=3 independent experiments, $^{**}p<0.01$; Student's t-test. **f,** Immunofluorescence staining of RAD51 focus formation in Kelly NB cells treated with THZ531 (400 nM) or DMSO for 24 h prior to exposure to gamma radiation (IR, 8 Gy). Nuclei are stained with DAPI (scale bar, 100 μM). Quantification of staining (right) reported as mean ± SD of $RAD51^+$ cells, n=3 independent experiments. $^{**}p<0.01$, $^{***}p<0.001$; Student's t-test.

Fig. 2. CDK12/13 inhibition leads to an elongation defect that is gene-length dependent. a, Average metagene profiles of normalized TT-seq reads over gene bodies and extending -2 kb to +2 kb of all detected genes in cells treated with THZ531 400 nM for 2 h. Sense and antisense reads are depicted by solid and dashed lines respectively. **b,** Average metagene profile depicting the **c**hange (red) and rate of change (blue) in TT-seq read densities in regions flanking the TSS (-0.5 kb to +1 kb) in cells treated with THZ531. **c,** Scatter plot showing log2 fold-changes in gene expression vs. gene length in log2 scale for each protein coding gene in cells treated as in panel **a** (R^2=0,12, p=2e-277, F-test, Spearman correlation coefficient = -0.42). Differentially expressed genes are indicated (FDR < 0.1 and |log2 FC| > 1). **d**, Average metagene profiles for protein-coding genes (as in panel **a**)

stratified according to quartiles of gene length distribution and for very short genes. Sense and antisense reads are depicted by solid and dashed lines respectively. **e,** Gene ontology (GO) enrichment analysis of long genes (>64.5 kb) (top); TT-seq tracks of nascent RNA expression at the *PCF11* locus in NB cells treated with DMSO or THZ531 as in panel **a** (bottom). **f,** GO enrichment analysis of very short genes (< 3.4 kb) (top); TT-seq tracks at the *HIST1H3A/HIST1H4A* loci (bottom).

Fig. 3. CDK12 inhibition leads to PCPA of long genes. a, Average metagene profiles of normalized poly(A) 3′-seq reads at the transcription end sites (TES) (– 1 kb to + 4 kb) of all long genes (> 64.5 kb) (left), and short genes (RD histone genes) (right). **b,** Average metagene profiles of normalized poly(A) 3'-seq reads over gene bodies and extending -2 kb to +2 kb of all detected genes in cells treated with THZ531 400 nM for 2 and 6 h. Sense and antisense reads are depicted by solid and dashed lines, respectively. **c,** Histograms showing the genomic distributions and rankings of the top 5000 poly(A) 3'-seq peaks in DMSO- and THZ531-treated cells (400 nM, 6 h). The poly(A) 3' peaks were binned according to the depicted genomic regions and their intensities (x-axis). **d,** Bar plot indicating the number of protein-coding genes that underwent premature cleavage and polyadenylation (PCPA) with THZ531. The expanded window on the right shows the genomic distribution of the identified intronic poly(A) sites. **e,** Average metagene profiles of normalized poly(A) 3′-seq reads at the TSS (- 1 kb to + 10 kb) for all detected genes in Kelly WT (left) and Kelly E9R (right) cells. Changes (insets) in read density between DMSO- and THZ531 (200 nM, 6h)-treated Kelly WT (p = 5.8e-41) and Kelly E9R (p = 0.11) cells; comparisons between groups by Wilcoxon rank-sum test. **f,** Density plot of odds-ratios of poly(A) site usage (intronic vs 3' UTR) for genes in Kelly WT and E9R cells (p = 0, Kolmogorov-Smirnov test,).

Fig. 4. CDK12/13 inhibition results in minimal splicing alterations. a, Diagrammatic representation (left) and bar plot of splicing events (right) observed in TT-seq analysis of NB cells treated with THZ531 (400 nM) for 2h. **b,** Scatterplot of intron retention index (IR index) vs. the ratio of exon and intron lengths in log2 scale. Genes with an IR index >1 or ≤1 display intron retention and loss respectively (adjusted $p < 0.05$, Fisher's exact test). **c,** Box plot illustrating the length distributions of genes that display intron loss or retention. **d,** Density plots illustrating the contributions of the proximal (first intron/exon) and distal (last intron/exon) gene regions in

calculation of the IR index. Comparison of IR indices distribution between proximal and distal intron/exon pairs ($p = 0$, Kolmogorov-Smirnov test).

Fig. 5. Gene length and a lower U1/PAS ratio predispose DDR genes to PCPA. a, GO enrichment analysis of the 809 genes that underwent PCPA (FDR < 0.01) based on TT-seq analysis of cells treated with THZ531 (400 nM for 2 h). **b,** Box plots and bar plots showing the distribution and numbers of PCPA and DDR genes in the different gene-length categories established in Fig. **2d** (****$p < 0.0001$, **$p < 0.01$, Fisher's exact test). **c,** TT-seq and poly(A) 3'-seq tracks at the *BLM* DDR gene locus depicting the loss of annotated terminal polyadenylation signal and the presence of early termination due to PCPA in cells treated with THZ531 as in panel **a**. **d,** Number of intronic poly(A) sites as a function of transcript length. A polynomial regression curve is plotted for all genes (black) and DDR genes only (red) ($p = 1.7e-13$, predicted vs. observed, Wilcoxon rank-sum test). **e,** Box plots comparing the indicated determinants of PCPA in all genes versus PCPA genes only and the proportion of DDR genes within the latter subset (see figure for *p* and *d* values; Wilcoxon rank-sum test & Cohen's *d* effect-size, respectively). **f,** Cumulative fraction plots showing the change in expression of PCPA ($p = 2.2e-16$, Kolmogorov-Smirnov test) and DDR ($p = 1.9e-14$, Kolmogorov-Smirnov test) transcripts relative to other transcripts following THZ531 treatment as in panel **a**. THZ, THZ531.

Fig. 6. CDK12/13 phosphorylates RNA processing proteins. a, Volcano plot of proteome-wide changes in phosphorylation site occupancy identified through SILAC analysis of NB cells treated with THZ531, 400 nM for 2 h. Expanded box shows selected co-transcriptional RNA processing proteins. **b,** GO terms for candidate CDK12/13 substrates. **c**, In vitro kinase assays of CDK12/CycK (red)- and CDK13/CycK (green)-mediated phosphorylation of CDC5L (aa 370-505) and GST-SF3B1 (aa 113-462) at the indicated time points. A negative control measurement without kinase is shown in blue. Radioactive kinase reactions were performed with 0.2 µM CDK12/CycK or CDK13/CycK and 50 µM substrate protein, respectively. Data are reported as mean ± SD, n=3. **d,** Dose-response curves of THZ531 incubated with recombinant CDK12 (left) and CDK13 (right) protein and CDC5L (aa 370-505) and GST-SF3B1 (aa 113-462). Radioactive kinase reactions were performed after 30 min preincubation with increasing concentrations of THZ531. For all incubation time series, the counts per minute of

the kinase activity measurements were normalized to the relative ^{32}P transfer. Data are reported as mean ± SD; n=3. **e**, Immunofluorescence imaging of nuclear speckles stained with an antibody against the splicing factor SRSF2 (anti-SC35) in NB cells treated with THZ531 (400nM), pladienolide B (100nM) or DMSO for 6 h. Nuclei are stained with DAPI. Scale bar, 100 μM. **f**, Model of CDK12 as a regulator of pre-mRNA processing. CDK12 phosphorylates and thus likely stimulates the orchestrated action of RNA Pol II CTD and RNA processing proteins. CDK12 inhibition leads to a gene-length-dependent elongation defect associated with early termination through premature cleavage and polyadenylation (PCPA). Especially vulnerable to PCPA are long genes with a lower ratio of U1 snRNP binding to poly(A) sites, which include many of those involved in the DDR. Among short genes, including genes that normally terminate through stem-loop binding, CDK12 inhibition increases intron retention and leads to longer polyadenylated transcripts.

Supplementary Figure Legends

Supplementary Fig. 1. CDK12/13 inhibition impairs viability of NB cells. a, IC_{50} values of THZ531 in NB vs. fibroblast (NIH-3T3, IMR-90, BJ) cells at 72 h. **b,** Analysis of target engagement in NB cells following THZ531 treatment. Cells were treated with THZ531 or DMSO for 6 h at the indicated concentrations and cell lysates incubated with 1μM of biotinylated THZ531 (bio-THZ531) overnight. **c**, Flow cytometric analysis of Annexin V (AV) staining in Kelly and IMR-32 NB cells treated with 400 nM THZ531 for the indicated times. Data represent mean ± SD; n=3, *$*p<0.05$* (Student's t-test). **d,** Western blot analysis of cleaved PARP in Kelly, IMR-32 and NGP cells following treatment with THZ531 at the indicated doses and times. **e,** Cell-cycle analysis of NB cells exposed to THZ531 (400 nM for 2, 6 and 24 h), by flow cytometry with propidium iodide (PI) staining. Results are representative of three replicate experiments. **f**, Dose-response curves for *MYCN*-amplified (left) and non-amplified (right) human NB cells. Cells were treated with increasing concentrations of THZ531 alone or in combination with the ABCB1 inhibitor, tariquidar (125 nM) for 72 h. Percent cell viability relative to DMSO-treated cells is shown. Data represent mean ± SD; n=3.

Supplementary Fig. 2. CDK12/13 inhibition with THZ531 preferentially affects DNA damage response genes. a, Heat map of gene expression values in NB cells treated with THZ531 (400 nM for 6 h) vs. DMSO. **b,**

Gene set enrichment analysis (GSEA) of downregulated genes in NB cells treated as in panel **a**. **c**, Waterfall plot of log2 fold-changes in gene expression in Kelly NB cells treated as in panel **a**; selected DDR genes are highlighted. **d**, qRT-PCR (left) and immunoblot (right) analyses of selected DDR gene expression in NB cells treated with THZ531 at the indicated concentrations for 6 h. **e,** Immunoblot analysis (left) of CDK12 and CDK13 expression following shRNA knockdown in NB cells. β-actin was used as a loading control. qRT-PCR (right) of selected DDR expression in NB cells expressing either a control shRNA or two individual shRNAs targeting CDK12 or CDK13. The qRT-PCR data in **d** and **e** were normalized to GAPDH and presented as mean ± SD; n=3**.** **f.** Flow cytometry analysis of γ-H2AX staining in IMR-32 NB cells treated with 400 nM THZ531 for the indicated time points (left). Quantification of staining (right) presented as the mean ± SD; n = 3 (right). ***p*<0.01, Student's t-test.

Supplementary Fig. 3. CDK12/13 inhibition results in genome-wide alterations in nascent RNA expression. **a,** Bar plot showing the genomic distribution of TT-seq reads for each replicate of the three conditions tested (DMSO and THZ531, 400 nM for 30 min and 2 h). **b,** Volcano plot representation of differentially expressed genes following treatment with THZ531 for 30 min (left) and 2 h (right). The fold changes are represented in log2 scale (*x*-axis), and the $-\log_{10}$ *q*-value depicted on the *y*-axis (FDR < 0.1 and |log2 FC| >1). Bar plots depict the numbers of up- and downregulated genes for each aggregated gene group in cells treated with THZ531 for 30 min and 2 h. **c,** Waterfall (upper) and density (lower) plots of log2 fold-changes in gene expression in IMR-32 NB cells treated with 400 nM THZ531 for 2 h; the aggregated set of DDR genes is highlighted in red. **d**, Distribution of average simulated changes in gene groups of similar size to the aggregated set of DDR genes. The z-score for the DDR gene set is depicted in red. **e,** qRT-PCR analysis of nascent RNA expression at selected regions of the *BRCA1* gene in NB cells treated with THZ531 at the indicated concentrations for 6 h. **f,** GO enrichment analysis of the top 400 downregulated genes after THZ531 (400 nM) treatment for 2 h.

Supplementary Fig. 4. Noncoding genes and very short genes do not exhibit a length-dependent elongation defect with CDK12/13 inhibition. a, Average metagene profile depicting the **c**hange (red) and rate of change (blue) in TT-seq read densities in regions flanking the TSS (-1 kb to +10 kb) in cells treated with THZ531 (400

nM) for 2 h. **b,** Average metagene profiles of normalized TT-seq reads of long and short noncoding genes (top) in cells treated with THZ531 (400 nM) for 2 h. Associated scatterplots of log2 fold changes in gene expression versus gene length in log2 scales for each gene (bottom)**.** Sense and antisense reads are depicted by solid and dashed lines, respectively. **c,** Heat map depicting the association between short genes and increased 3′UTR length. Blue and red colors represent negative and positive odds ratios, respectively (Fisher's exact test). **d**, Average metagene profiles of normalized TT-seq reads of RD histone genes. **e,** TT-seq and poly(A) 3′-seq tracks at the *MYCN* locus showing 3′ UTR extension with usage of distal 3′ poly(A) sites upon treatment with THZ531. Enlarged box shows multiple poly(A) 3′-seq peaks downstream of the last dominant peak. **f,** qRT-PCR analysis of total RNA (top) and nascent RNA expression (bottom) at the indicated regions (A, B, C, D) of the *MYCN* gene in NB cells treated with THZ531 at the indicated concentrations for 6 h. Data normalized to GAPDH are presented as mean ± SD; n=3**.**

Supplementary Fig. 5. CDK12 depletion induces a higher usage of intronic polyadenylation sites than CDK13 depletion. a, Bar plots depicting the frequency of retrieved polyadenylation site (PAS) motifs 100 bp upstream of the poly(A) 3′-seq peaks. **b,** Scatterplot of poly(A) 3'-seq reads for all detected peaks in DMSO (x-axis) and THZ531-treated (y-axis) cells. Peaks with a 2-fold increase or decrease were considered to be THZ531- or DMSO-specific respectively. **c,** Pie charts depicting the genomic distributions of poly(A) 3'-seq peaks from panel **b,** in DMSO- and THZ531-treated cells. **d,** Average metagene profiles of rescaled (1-100) and normalized TT-seq and poly(A) 3′-seq reads at the TSS (- 1 kb to + 10 kb) for all PCPA genes in cells treated with DMSO or THZ531. **e,** Average metagene profiles of normalized poly(A) 3′-seq reads at the TSS (- 1 kb to + 10 kb) in IMR-32 NB cells expressing shRNAs against CDK12 and CDK13. Cells expressing a shRNA against GFP were used as controls. *Inset*, box plots depicting the change in distribution values over above-mentioned range (-1 kb to +10 kb) (shCDK12 vs. shGFP, ****p = 5.3e-13; shCDK13 vs. shGFP, **p = 0.01; shCDK12 vs. shCDK13, ****p = 4.8e-6; Wilcoxon rank-sum test). **f,** Density plots of odds ratios of poly(A) site usage (intronic vs 3′UTR) in the cells described in panel **e** (shCDK12 vs. shCDK13, p = 0; Kolmogorov-Smirnov test). **g,** Scatterplot of the log2 fold changes (vs. control) in gene expression versus gene length in log2 scale for each protein-coding gene in

IMR-32 and Kelly NB cells under the following conditions: IMR-32 cells treated with THZ531, 400 nM (left), Kelly WT and Kelly E9R cells treated with THZ531, 200 nM (middle), for the indicated times; IMR-32 cells expressing shRNAs against CDK12 and CDK13 (right). For each condition, a generalized additive model (GAM) smoothing curve depicting the general trend in expression change is shown underneath each scatterplot.

Supplementary Fig. 6. CDK12 depletion results in a gene length-dependent loss of expression. a,.Density plot showing the distribution of gene length among all genes vs. that among PCPA genes (p = 2e-11, Wilcoxon rank-sum test). **b,** Table depicting the identified DDR genes within the PCPA gene group and stratified according to the gene length groups established in Fig. **2d**. **c,** TT-seq and poly(A) 3'-seq tracks at the *BARD1* locus showing loss of annotated terminal polyadenylation signal and early termination due to PCPA upon treatment with THZ531 (400 nM, 6 h) (left). Agarose gel electrophoresis of 3'-RACE products of *BARD1* in NB cells treated with THZ531 or DMSO (right). PCR products were validated by DNA sequencing. Arrow indicates the short isoform.

Supplementary Fig. 7. CDK12 regulates the processing of DDR gene transcripts. a, Comparison of PCPA genes identified in Oh *et al. (2017)*[42] after U1 inhibition with a U1 antisense morpholino oligonucleotide (AMO) at 4 h (p = 7.52e-52 and odds ratio = 3.5) and 8 h (p = 2e-38 & odds ratio = 2.7) with PCPA genes identified in TT-seq analysis of cells treated with THZ531 400 nM for 2 h (all comparisons by Fisher's exact test). **b,** Density plots of odds ratios of poly(A) site usage (intronic vs 3'UTR) in all genes (blue) vs. DDR genes only (red) in Kelly WT and Kelly E9R cells treated with THZ531 200 nM for 6 h (p = 5.6e-4 and p = 0.4, respectively; Kolmogorov-Smirnov test). **c**, GO enrichment of genes that display increased intronic poly(A) usage in Kelly WT cells described in panel **b**. (FDR < 0.01). **d,** Cumulative fraction plots illustrating changes in the expression of DDR genes in Kelly WT and Kelly E9R cells (****p = 3.3e-16, Kolmogorov-Smirnov test). **e,** Heat map of log2 fold changes in gene expression values in Kelly WT and Kelly E9R cells treated with THZ531 (400 nM, 6 h) vs. DMSO.

Supplementary Fig. 8. CDK12/13 inhibition affects phosphorylation of RNA processing genes. a, Consensus sequences derived from the phosphopeptides showing decreased phosphorylation in cells treated with THZ531 (400nM, 2 h) using WebLogo 3 software, (http://weblogo.threeplusone.com/create.cgi). The phosphorylation site detected in SILAC analysis is labeled as position 0 in the plot. **b**, Protein-protein interaction network of the candidate phosphorylation substrates identified from SILAC analysis of NB cells as in panel **a**, using the STRING 10.5 database (http://string-db.org/). **c,** Coomassie blue staining of GST-tagged recombinant proteins used in the in vitro kinase assays. **d,** In vitro kinase assays of CDK12/CycK (red)- and CDK13/CycK (green)-mediated phosphorylation of GST-SPT6H (aa 1323-1544) at the indicated time points. A negative control measurement without kinase is shown in blue. Radioactive kinase reactions were performed using 0.2 μM CDK12/CycK or CDK13/CycK and 50 μM substrate protein, respectively. Data are shown as mean ± SD, n=3. **e**, Mass spectrometry spectra generated using Xcalibur software (Thermo Fisher Scientific) of a specific SPT6H peptide incubated with CDK12/CycK. The phosphorylation site identified by mass spectrometry is indicated. **f,** Quantification of SRSF2-stained nuclear speckle size using ImageJ software in NB cells treated with THZ531 (400 nM), pladienolide B (100 nM) or DMSO for 6 hr. Data are represented as mean ± SD, n=3; ***p<0.001; Student's t-test.

References

1 Buratowski, S. The CTD code. *Nat Struct Biol* **10**, 679-680, doi:10.1038/nsb0903-679 (2003).

2 Bentley, D. L. Coupling mRNA processing with transcription in time and space. *Nat Rev Genet* **15**, 163-175, doi:10.1038/nrg3662 (2014).

3 Ho, C. K. & Shuman, S. Distinct roles for CTD Ser-2 and Ser-5 phosphorylation in the recruitment and allosteric activation of mammalian mRNA capping enzyme. *Mol Cell* **3**, 405-411 (1999).

4 Ramanathan, Y. *et al.* Three RNA polymerase II carboxyl-terminal domain kinases display distinct substrate preferences. *J Biol Chem* **276**, 10913-10920, doi:10.1074/jbc.M010975200 (2001).

5 Bartkowiak, B. *et al.* CDK12 is a transcription elongation-associated CTD kinase, the metazoan ortholog of yeast Ctk1. *Genes Dev* **24**, 2303-2316, doi:10.1101/gad.1968210 (2010).

6 Blazek, D. *et al.* The Cyclin K/Cdk12 complex maintains genomic stability via regulation of expression of DNA damage response genes. *Genes Dev* **25**, 2158-2172, doi:10.1101/gad.16962311 (2011).

7 Cheng, S. W. *et al.* Interaction of cyclin-dependent kinase 12/CrkRS with cyclin K1 is required for the phosphorylation of the C-terminal domain of RNA polymerase II. *Mol Cell Biol* **32**, 4691-4704, doi:10.1128/MCB.06267-11 (2012).

8 Ko, T. K., Kelly, E. & Pines, J. CrkRS: a novel conserved Cdc2-related protein kinase that colocalises with SC35 speckles. *J Cell Sci* **114**, 2591-2603 (2001).

9 Malumbres, M. Cyclin-dependent kinases. *Genome Biol* **15**, 122 (2014).

10 Liang, K. *et al.* Characterization of human cyclin-dependent kinase 12 (CDK12) and CDK13 complexes in C-terminal domain phosphorylation, gene transcription, and RNA processing. *Mol Cell Biol* **35**, 928-938, doi:10.1128/MCB.01426-14 (2015).

11 Bajrami, I. *et al.* Genome-wide profiling of genetic synthetic lethality identifies CDK12 as a novel determinant of PARP1/2 inhibitor sensitivity. *Cancer Res* **74**, 287-297, doi:10.1158/0008-5472.CAN-13-2541 (2014).

12 Joshi, P. M., Sutor, S. L., Huntoon, C. J. & Karnitz, L. M. Ovarian cancer-associated mutations disable catalytic activity of CDK12, a kinase that promotes homologous recombination repair and resistance to cisplatin and poly(ADP-ribose) polymerase inhibitors. *J Biol Chem* **289**, 9247-9253, doi:10.1074/jbc.M114.551143 (2014).

13 Johnson, S. F. *et al.* CDK12 Inhibition Reverses De Novo and Acquired PARP Inhibitor Resistance in BRCA Wild-Type and Mutated Models of Triple-Negative Breast Cancer. *Cell Rep* **17**, 2367-2381, doi:10.1016/j.celrep.2016.10.077 (2016).

14 Wu, Y. M. *et al.* Inactivation of CDK12 Delineates a Distinct Immunogenic Class of Advanced Prostate Cancer. *Cell* **173**, 1770-1782 e1714, doi:10.1016/j.cell.2018.04.034 (2018).

15 Zhang, T. *et al.* Covalent targeting of remote cysteine residues to develop CDK12 and CDK13 inhibitors. *Nat Chem Biol* **12**, 876-884, doi:10.1038/nchembio.2166 (2016).

16 Iniguez, A. B. *et al.* EWS/FLI Confers Tumor Cell Synthetic Lethality to CDK12 Inhibition in Ewing Sarcoma. *Cancer Cell* **33**, 202-216 e206, doi:10.1016/j.ccell.2017.12.009 (2018).

17 Chipumuro, E. *et al.* CDK7 inhibition suppresses super-enhancer-linked oncogenic transcription in MYCN-driven cancer. *Cell* **159**, 1126-1139, doi:10.1016/j.cell.2014.10.024 (2014).

18 Schleiermacher, G. *et al.* Segmental chromosomal alterations have prognostic impact in neuroblastoma: a report from the INRG project. *Br J Cancer* **107**, 1418-1422, doi:10.1038/bjc.2012.375 (2012).

19 Molenaar, J. J. *et al.* Sequencing of neuroblastoma identifies chromothripsis and defects in neuritogenesis genes. *Nature* **483**, 589-593, doi:10.1038/nature10910 (2012).

20 Pugh, T. J. *et al.* The genetic landscape of high-risk neuroblastoma. *Nat Genet* **45**, 279-284, doi:10.1038/ng.2529 (2013).

21 Zhang, H. *et al.* Targeting CDK9 Reactivates Epigenetically Silenced Genes in Cancer. *Cell*, doi:doi.org/10.1016/j.cell.2018.09.051 (2018).

22 Gao, Y. *et al.* Overcoming Resistance to the THZ Series of Covalent Transcriptional CDK Inhibitors. *Cell Chem Biol* **25**, 135-142 e135, doi:10.1016/j.chembiol.2017.11.007 (2018).

23 Martin, C. *et al.* The molecular interaction of the high affinity reversal agent XR9576 with P-glycoprotein. *Br J Pharmacol* **128**, 403-411, doi:10.1038/sj.bjp.0702807 (1999).

24 Bartkowiak, B. & Greenleaf, A. L. Expression, purification, and identification of associated proteins of the full-length hCDK12/CyclinK complex. *J Biol Chem* **290**, 1786-1795, doi:10.1074/jbc.M114.612226 (2015).

25 Harlen, K. M. *et al.* Comprehensive RNA Polymerase II Interactomes Reveal Distinct and Varied Roles for Each Phospho-CTD Residue. *Cell Rep* **15**, 2147-2158, doi:10.1016/j.celrep.2016.05.010 (2016).

26 Hintermair, C. *et al.* Threonine-4 of mammalian RNA polymerase II CTD is targeted by Polo-like kinase 3 and required for transcriptional elongation. *EMBO J* **31**, 2784-2797, doi:10.1038/emboj.2012.123 (2012).

27 Hsin, J. P., Sheth, A. & Manley, J. L. RNAP II CTD phosphorylated on threonine-4 is required for histone mRNA 3' end processing. *Science* **334**, 683-686, doi:10.1126/science.1206034 (2011).

28 Kwiatkowski, N. *et al.* Targeting transcription regulation in cancer with a covalent CDK7 inhibitor. *Nature* **511**, 616-620, doi:10.1038/nature13393 (2014).

29 Prakash, R., Zhang, Y., Feng, W. & Jasin, M. Homologous recombination and human health: the roles of BRCA1, BRCA2, and associated proteins. *Cold Spring Harb Perspect Biol* **7**, a016600, doi:10.1101/cshperspect.a016600 (2015).

30 Ekumi, K. M. *et al.* Ovarian carcinoma CDK12 mutations misregulate expression of DNA repair genes via deficient formation and function of the Cdk12/CycK complex. *Nucleic Acids Res* **43**, 2575-2589, doi:10.1093/nar/gkv101 (2015).

31 Eifler, T. T. *et al.* Cyclin-dependent kinase 12 increases 3' end processing of growth factor-induced c-FOS transcripts. *Mol Cell Biol* **35**, 468-478, doi:10.1128/MCB.01157-14 (2015).

32 Tien, J. F. *et al.* CDK12 regulates alternative last exon mRNA splicing and promotes breast cancer cell invasion. *Nucleic Acids Res* **45**, 6698-6716, doi:10.1093/nar/gkx187 (2017).

33 Schwalb, B. *et al.* TT-seq maps the human transient transcriptome. *Science* **352**, 1225-1228, doi:10.1126/science.aad9841 (2016).

34 Shao, W. & Zeitlinger, J. Paused RNA polymerase II inhibits new transcriptional initiation. *Nat Genet* **49**, 1045-1051, doi:10.1038/ng.3867 (2017).

35 Gressel, S. *et al.* CDK9-dependent RNA polymerase II pausing controls transcription initiation. *Elife* **6**, doi:10.7554/eLife.29736 (2017).

36 Blazek, D. The cyclin K/Cdk12 complex: an emerging new player in the maintenance of genome stability. *Cell Cycle* **11**, 1049-1050, doi:10.4161/cc.11.6.19678 (2012).

37 Harris, M. E. *et al.* Regulation of histone mRNA in the unperturbed cell cycle: evidence suggesting control at two posttranscriptional steps. *Mol Cell Biol* **11**, 2416-2424 (1991).

38 Dominski, Z. & Marzluff, W. F. Formation of the 3' end of histone mRNA. *Gene* **239**, 1-14 (1999).

39 Tian, B., Pan, Z. & Lee, J. Y. Widespread mRNA polyadenylation events in introns indicate dynamic interplay between polyadenylation and splicing. *Genome Res* **17**, 156-165, doi:10.1101/gr.5532707 (2007).

40 Kaida, D. *et al.* U1 snRNP protects pre-mRNAs from premature cleavage and polyadenylation. *Nature* **468**, 664-668, doi:10.1038/nature09479 (2010).

41 Berg, M. G. *et al.* U1 snRNP determines mRNA length and regulates isoform expression. *Cell* **150**, 53-64, doi:10.1016/j.cell.2012.05.029 (2012).

42 Oh, J. M. *et al.* U1 snRNP telescripting regulates a size-function-stratified human genome. *Nat Struct Mol Biol* **24**, 993-999, doi:10.1038/nsmb.3473 (2017).

43 Chen, H. H., Wang, Y. C. & Fann, M. J. Identification and characterization of the CDK12/cyclin L1 complex involved in alternative splicing regulation. *Mol Cell Biol* **26**, 2736-2745, doi:10.1128/MCB.26.7.2736-2745.2006 (2006).

44 Rodrigues, F., Thuma, L. & Klambt, C. The regulation of glial-specific splicing of Neurexin IV requires HOW and Cdk12 activity. *Development* **139**, 1765-1776, doi:10.1242/dev.074070 (2012).

45 Heyn, P., Kalinka, A. T., Tomancak, P. & Neugebauer, K. M. Introns and gene expression: cellular constraints, transcriptional regulation, and evolutionary consequences. *Bioessays* **37**, 148-154, doi:10.1002/bies.201400138 (2015).

46 Zhang, J., Kuo, C. C. & Chen, L. GC content around splice sites affects splicing through pre-mRNA secondary structures. *BMC Genomics* **12**, 90, doi:10.1186/1471-2164-12-90 (2011).

47 Almada, A. E., Wu, X., Kriz, A. J., Burge, C. B. & Sharp, P. A. Promoter directionality is controlled by U1 snRNP and polyadenylation signals. *Nature* **499**, 360-363, doi:10.1038/nature12349 (2013).

48 Nigg, E. A. Cellular substrates of p34(cdc2) and its companion cyclin-dependent kinases. *Trends Cell Biol* **3**, 296-301 (1993).

49 Spritz, R. A. *et al.* The human U1-70K snRNP protein: cDNA cloning, chromosomal localization, expression, alternative splicing and RNA-binding. *Nucleic Acids Res* **15**, 10373-10391 (1987).

50 Grote, M. *et al.* Molecular architecture of the human Prp19/CDC5L complex. *Mol Cell Biol* **30**, 2105-2119, doi:10.1128/MCB.01505-09 (2010).

51 Mu, R. *et al.* Depletion of pre-mRNA splicing factor Cdc5L inhibits mitotic progression and triggers mitotic catastrophe. *Cell Death Dis* **5**, e1151, doi:10.1038/cddis.2014.117 (2014).

52 Wahl, M. C., Will, C. L. & Luhrmann, R. The spliceosome: design principles of a dynamic RNP machine. *Cell* **136**, 701-718, doi:10.1016/j.cell.2009.02.009 (2009).

53 Vos, S. M. *et al.* Structure of activated transcription complex Pol II-DSIF-PAF-SPT6. *Nature* **560**, 607-612, doi:10.1038/s41586-018-0440-4 (2018).

54 Zhong, X. Y., Wang, P., Han, J., Rosenfeld, M. G. & Fu, X. D. SR proteins in vertical integration of gene expression from transcription to RNA processing to translation. *Mol Cell* **35**, 1-10, doi:10.1016/j.molcel.2009.06.016 (2009).

55 Steurer, B. *et al.* Live-cell analysis of endogenous GFP-RPB1 uncovers rapid turnover of initiating and promoter-paused RNA Polymerase II. *Proc Natl Acad Sci U S A* **115**, E4368-E4376, doi:10.1073/pnas.1717920115 (2018).

56 Saldi, T., Fong, N. & Bentley, D. L. Transcription elongation rate affects nascent histone pre-mRNA folding and 3' end processing. *Genes Dev* **32**, 297-308, doi:10.1101/gad.310896.117 (2018).

57 Yurko, N. M. & Manley, J. L. The RNA polymerase II CTD "orphan" residues: Emerging insights into the functions of Tyr-1, Thr-4, and Ser-7. *Transcription* **9**, 30-40, doi:10.1080/21541264.2017.1338176 (2018).

58 Thapar, R. Structural basis for regulation of RNA-binding proteins by phosphorylation. *ACS Chem Biol* **10**, 652-666, doi:10.1021/cb500860x (2015).

59 Girard, C. *et al.* Post-transcriptional spliceosomes are retained in nuclear speckles until splicing completion. *Nat Commun* **3**, 994, doi:10.1038/ncomms1998 (2012).

60 Devany, E. *et al.* Intronic cleavage and polyadenylation regulates gene expression during DNA damage response through U1 snRNA. *Cell Discov* **2**, 16013, doi:10.1038/celldisc.2016.13 (2016).

61 Smart, A. C. *et al.* Intron retention is a source of neoepitopes in cancer. *Nat Biotechnol*, doi:10.1038/nbt.4239 (2018).

62 Lee, S. H. *et al.* Widespread intronic polyadenylation inactivates tumour suppressor genes in leukaemia. *Nature* **561**, 127-131, doi:10.1038/s41586-018-0465-8 (2018).

Figure 1

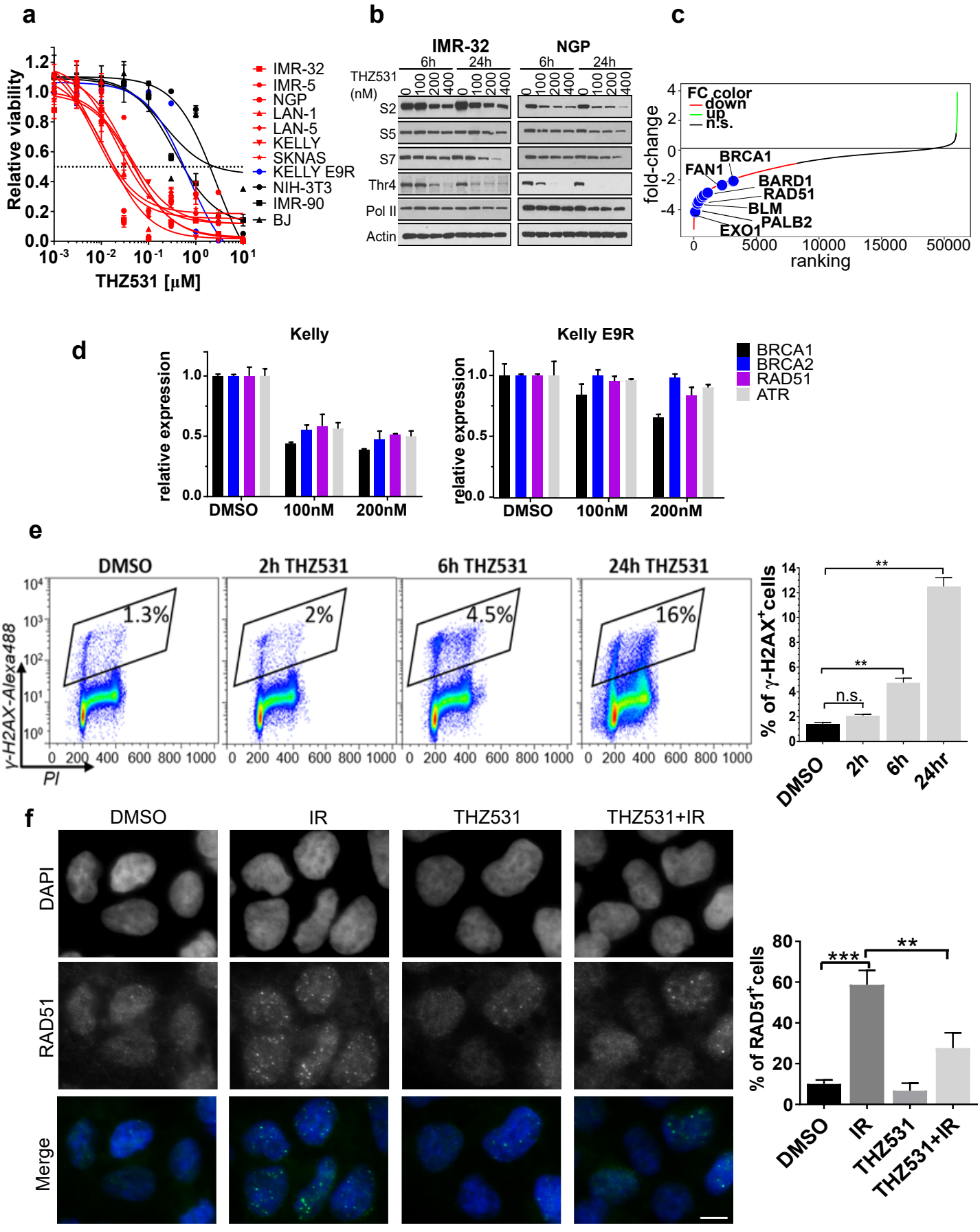


Figure 2

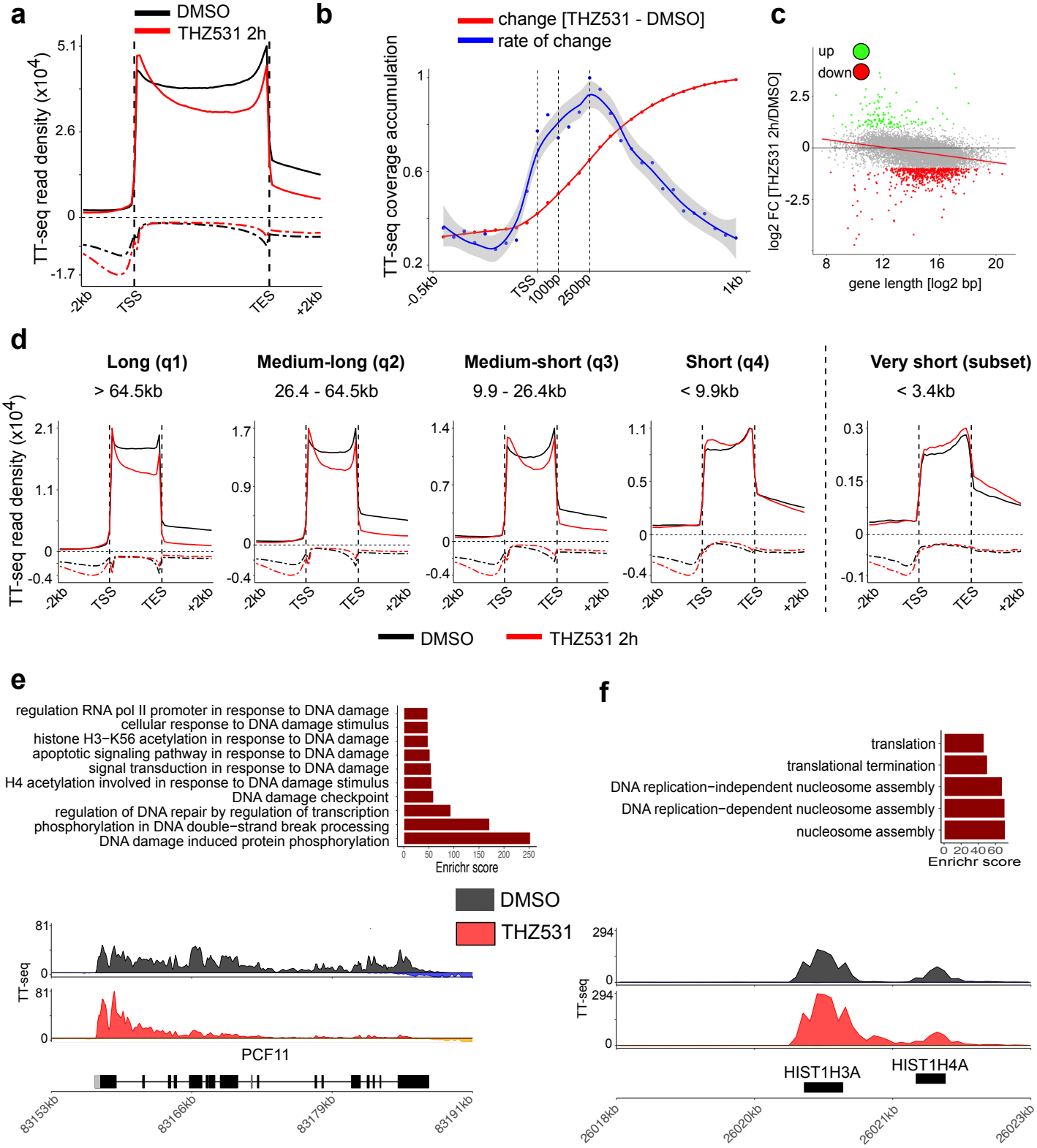


Figure 3

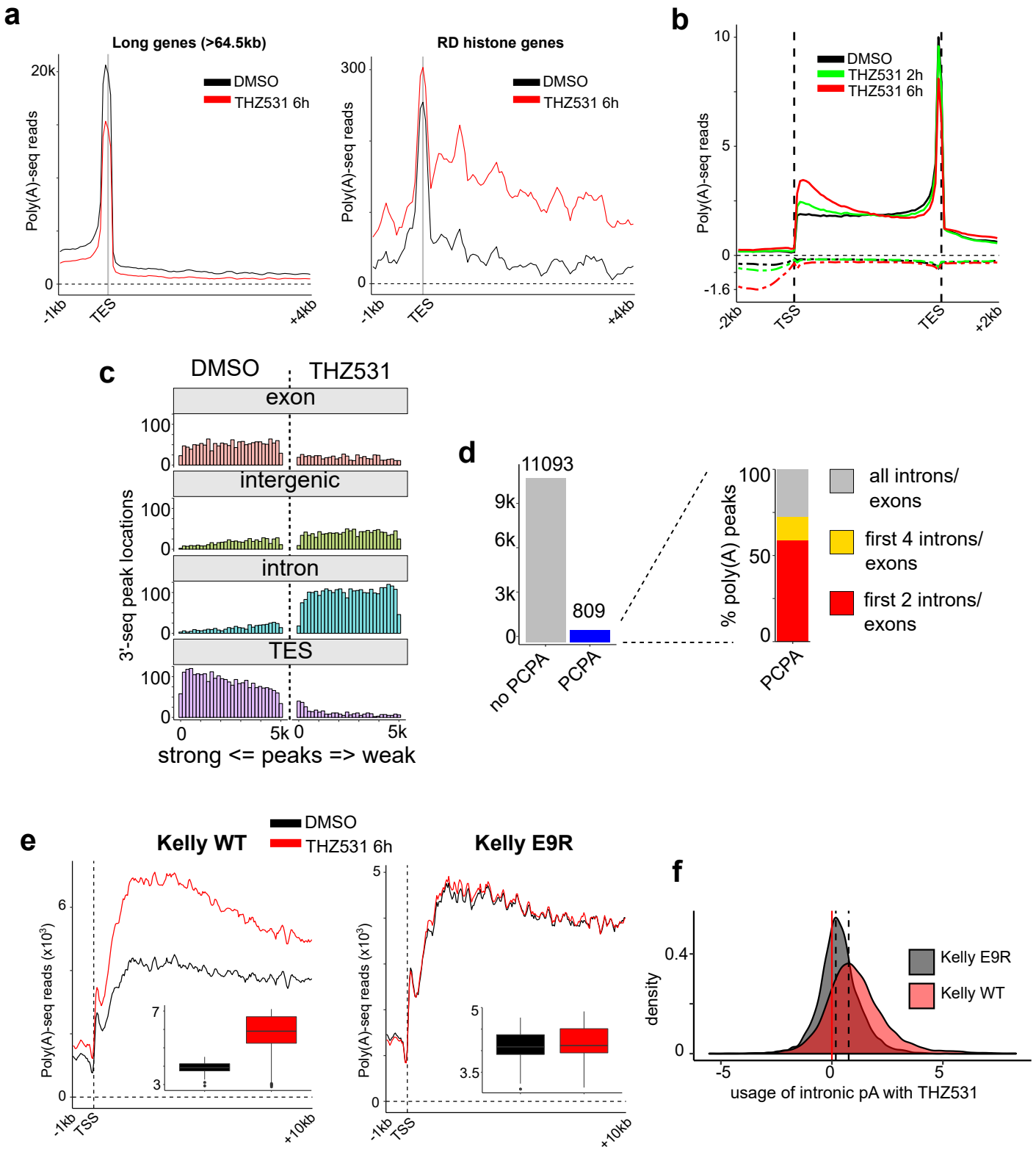


Figure 4

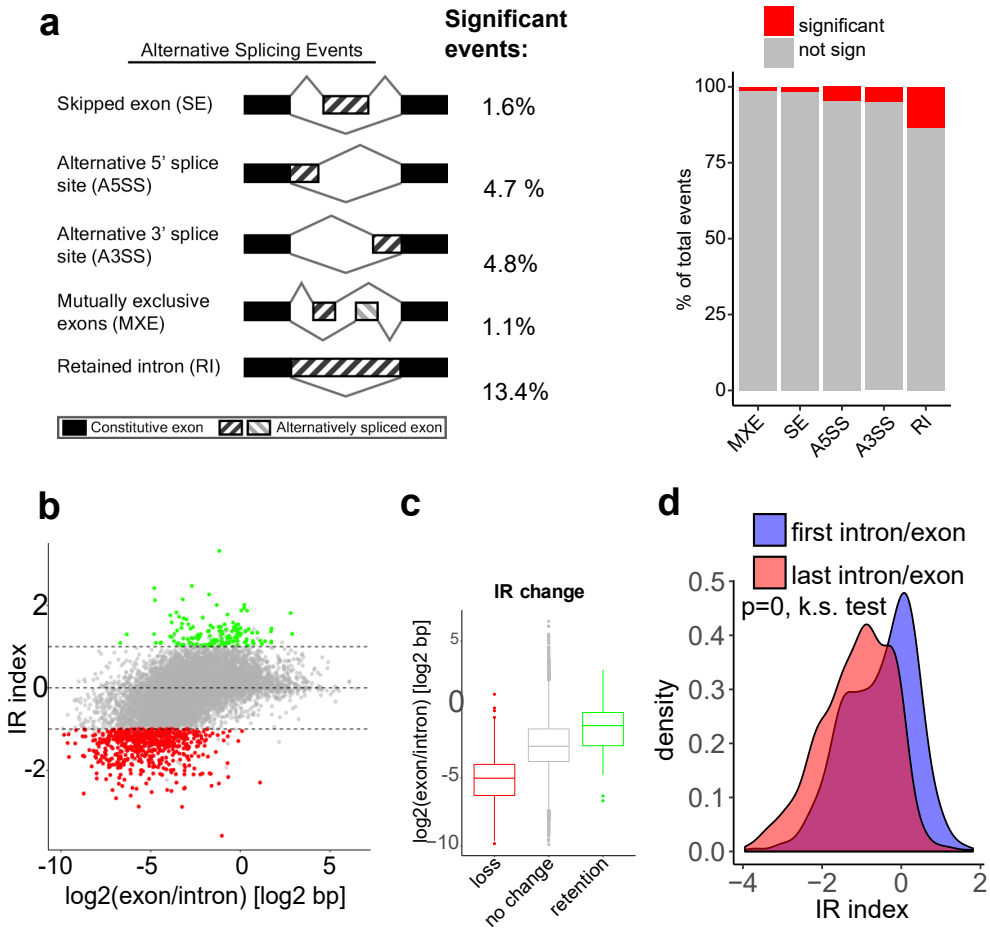


Figure 5

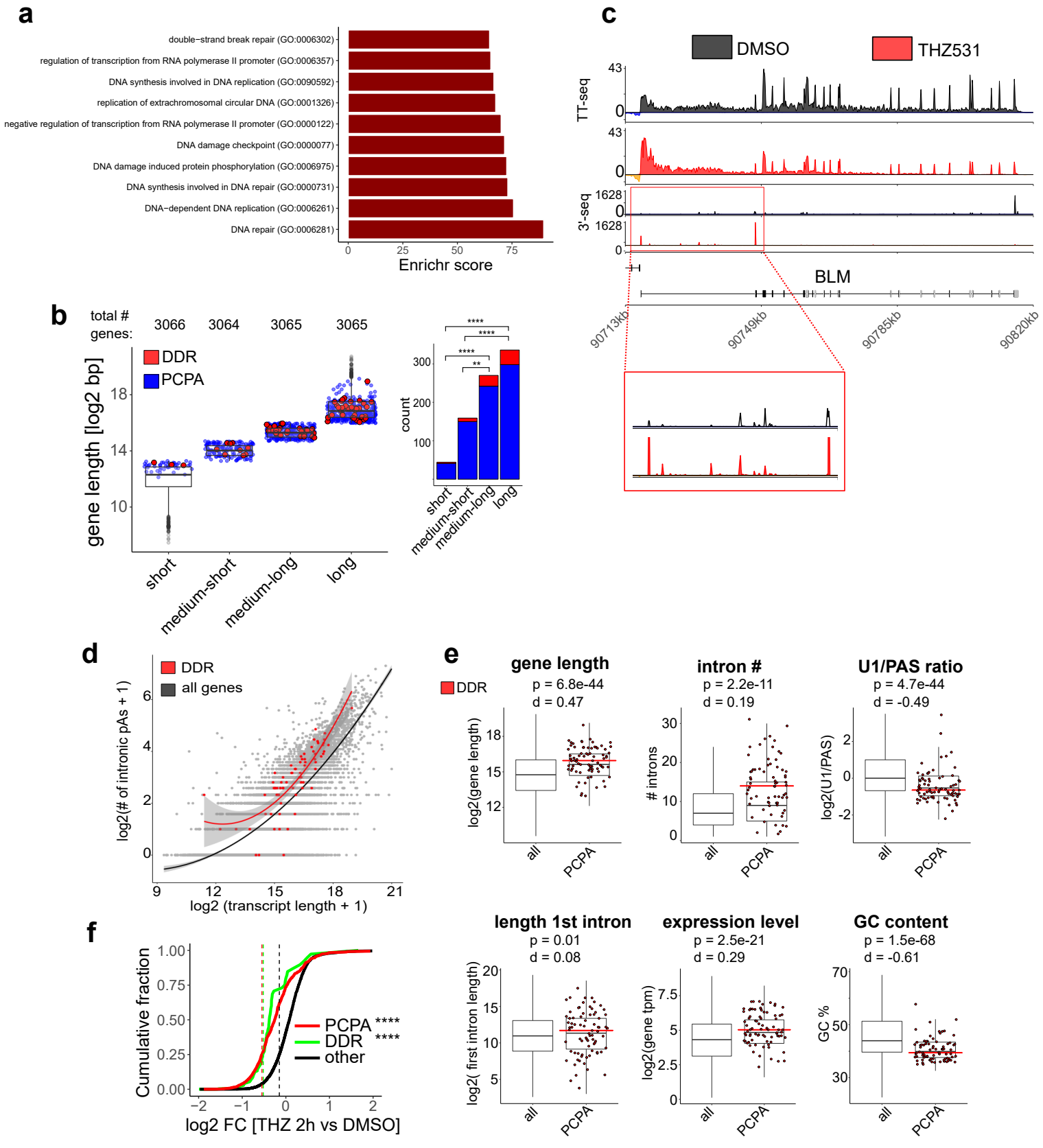
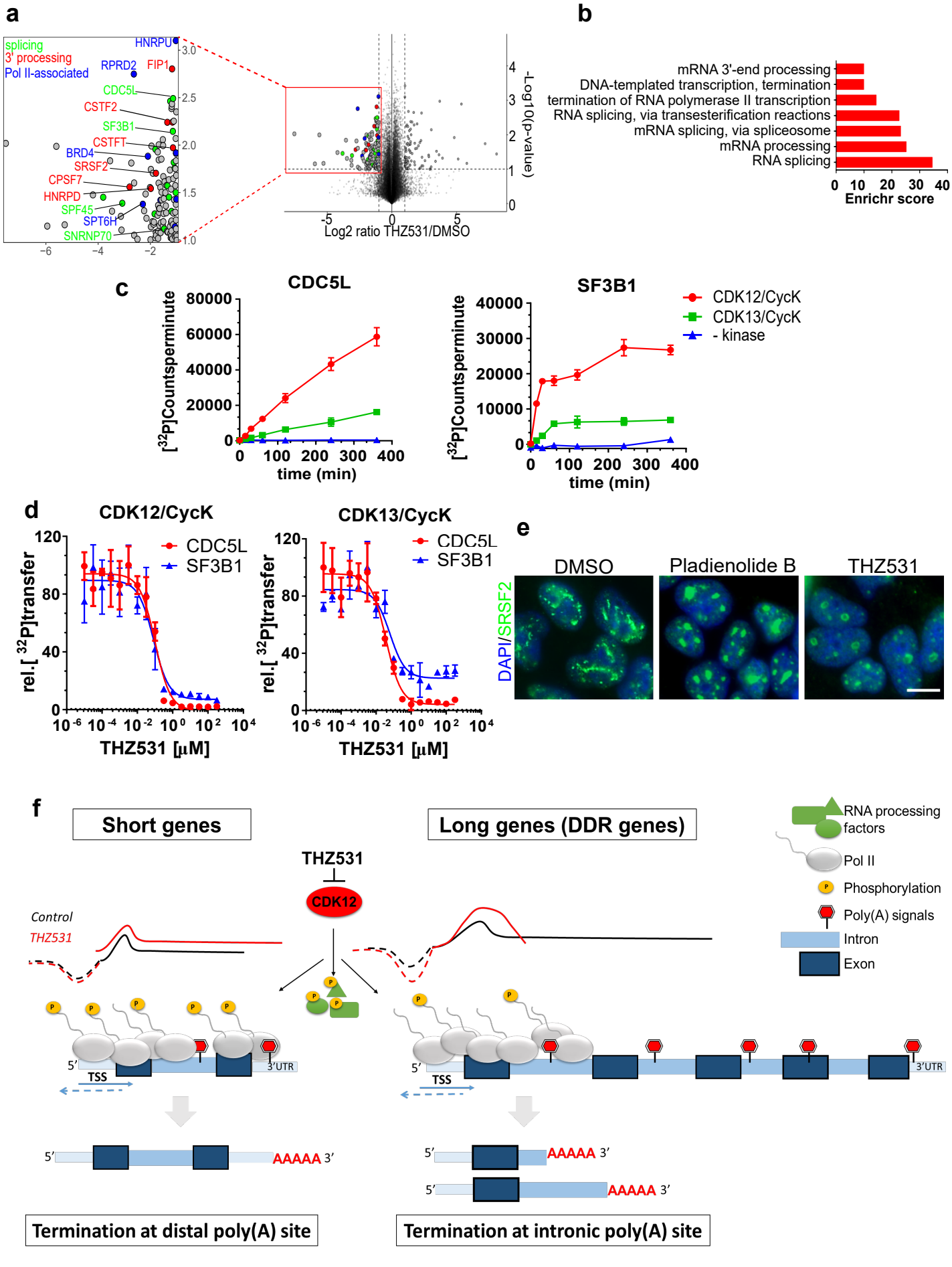


Figure 6

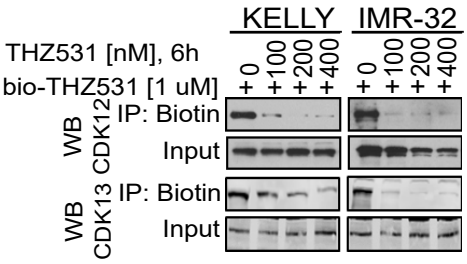


Supplementary Figure 1

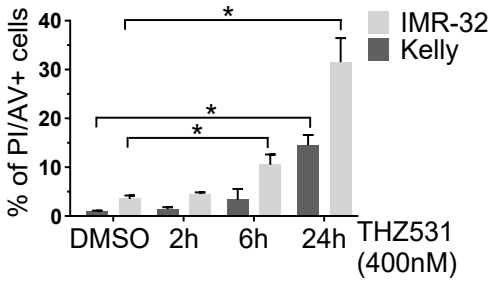
a

IC_{50} (nM) values of THZ531	
IMR-32	20
IMR-5	30
Kelly	60
NGP	35
LAN-1	30
LAN-5	35
SK-N-AS	85
Kelly E9R	400
IMR-90	800
NIH-3T3	1950
BJ	>2.000

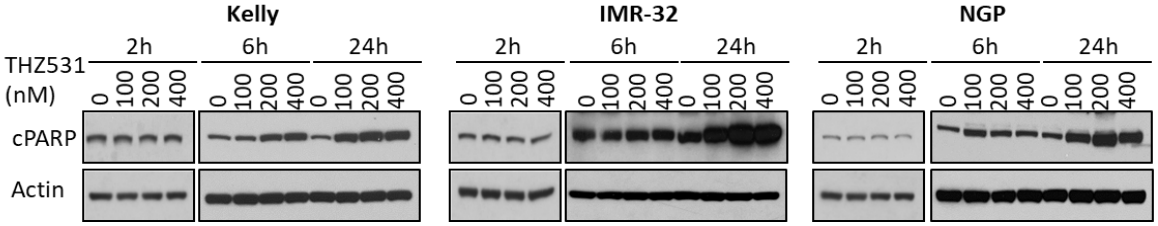
b



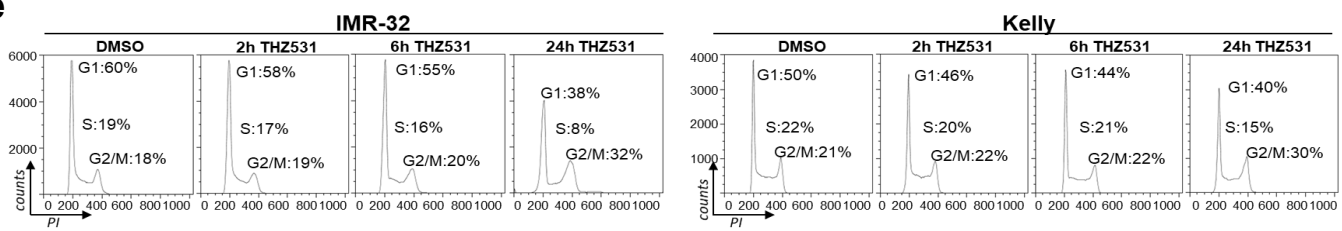
c



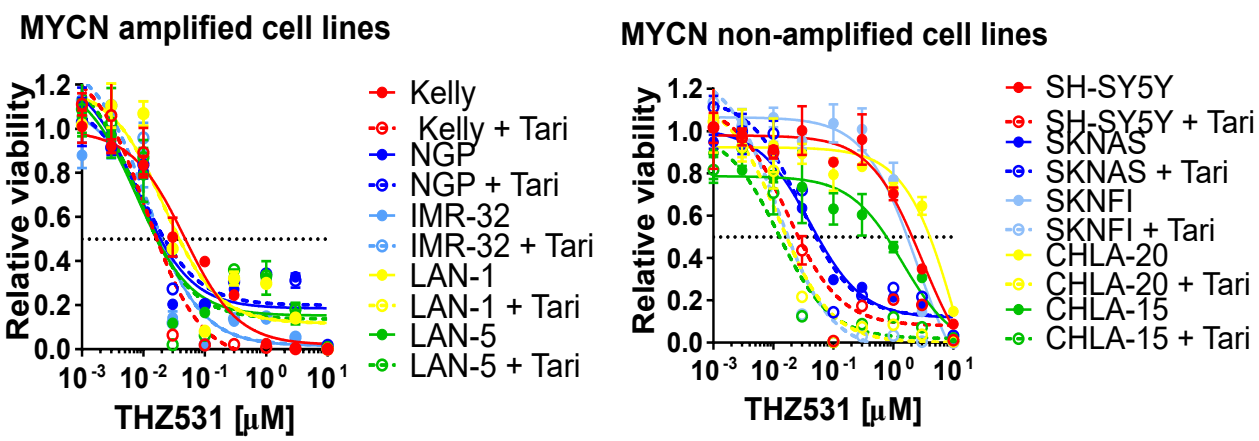
d



e



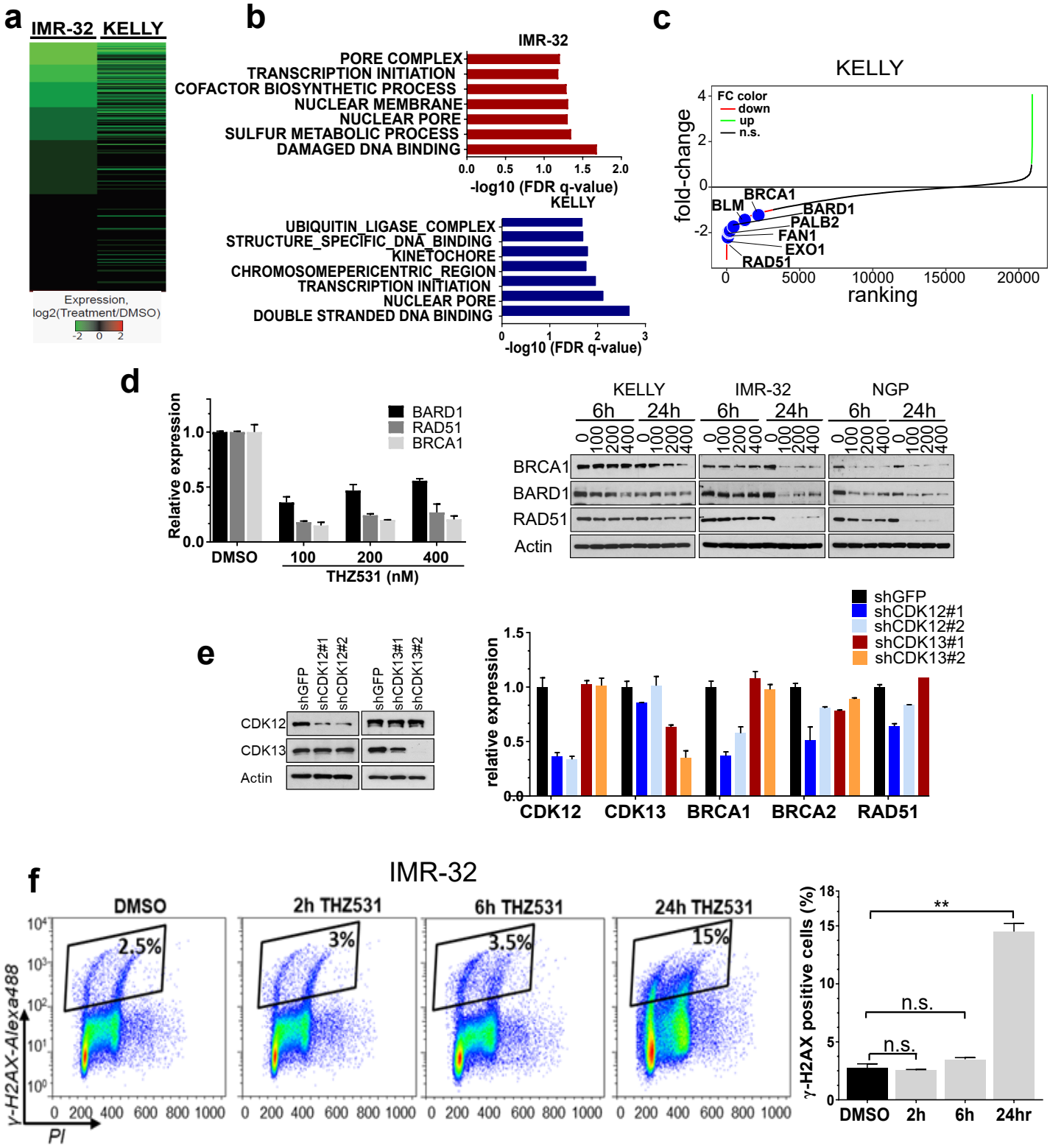
f



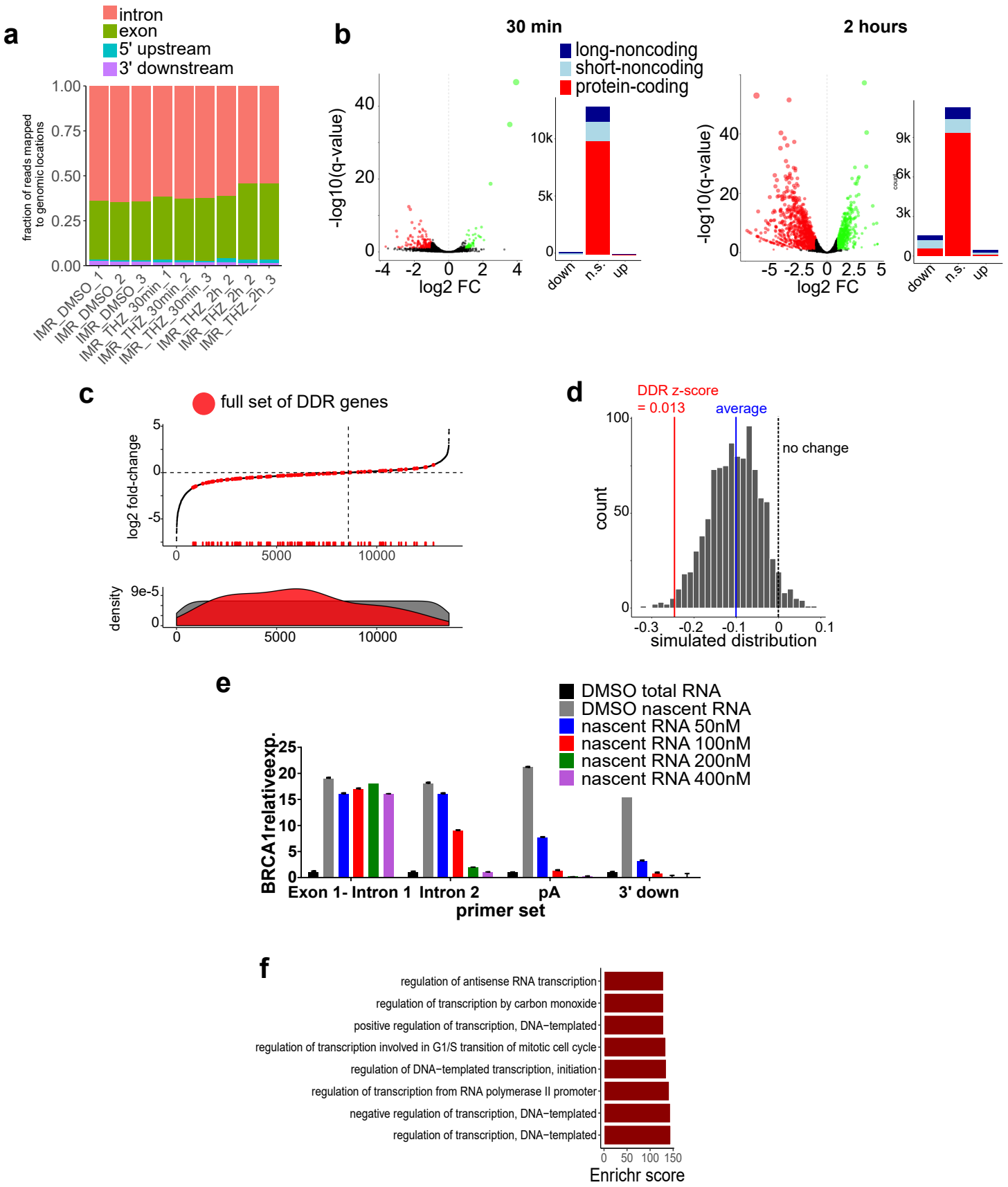
Cell line	IC_{50} (nM) THZ531 -	+Tariq
Kelly	60	15
NGP	30	29
IMR-32	20	18
LAN-1	30	29
LAN-5	35	15

Cell line	IC_{50} (nM) THZ531 -	+Tariq
SK-N-AS	85	60
SH-SY5Y	2000	30
SK-N-F1	2100	43
CHLA-15	600	15
CHLA-20	3500	20

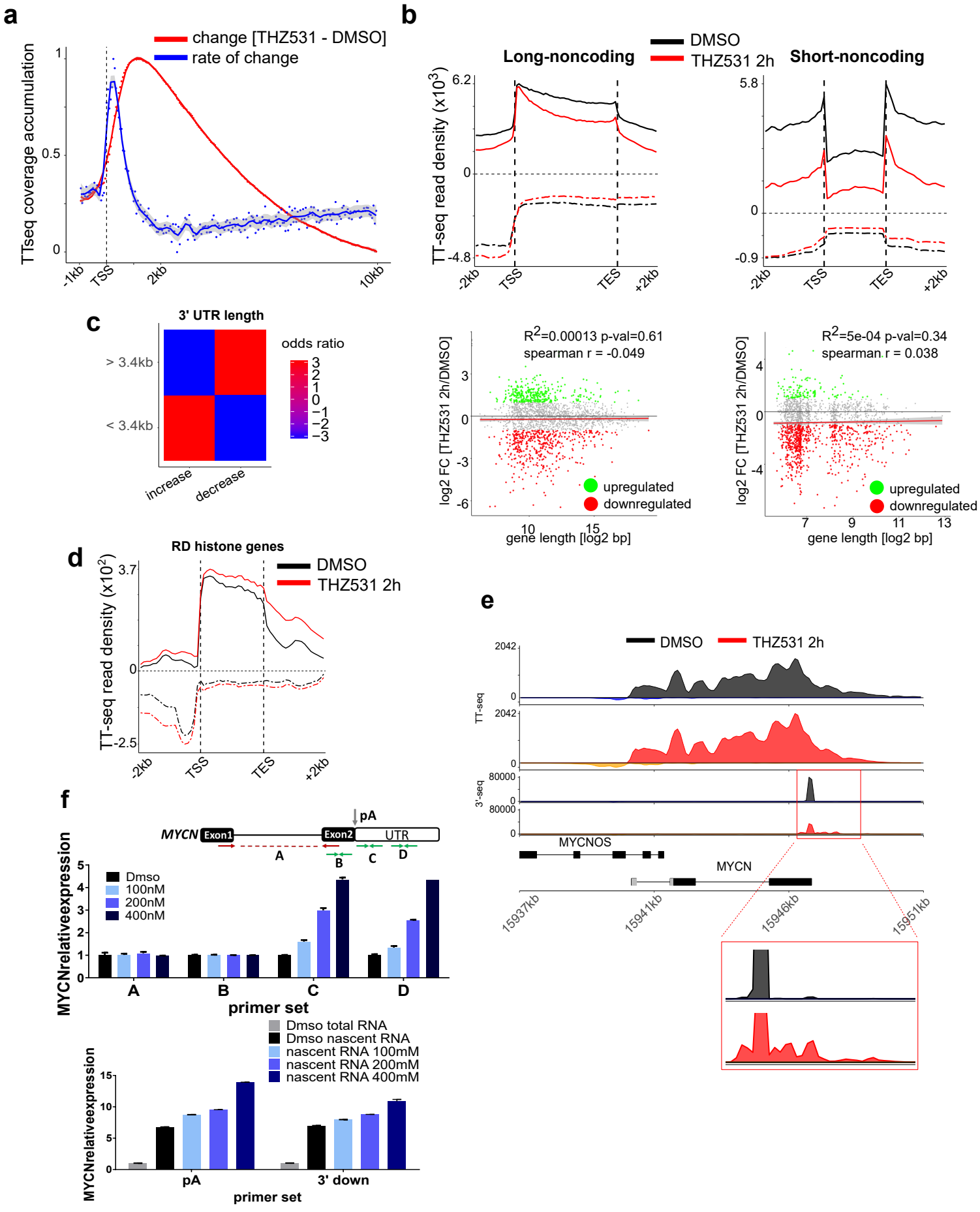
Supplementary Figure 2



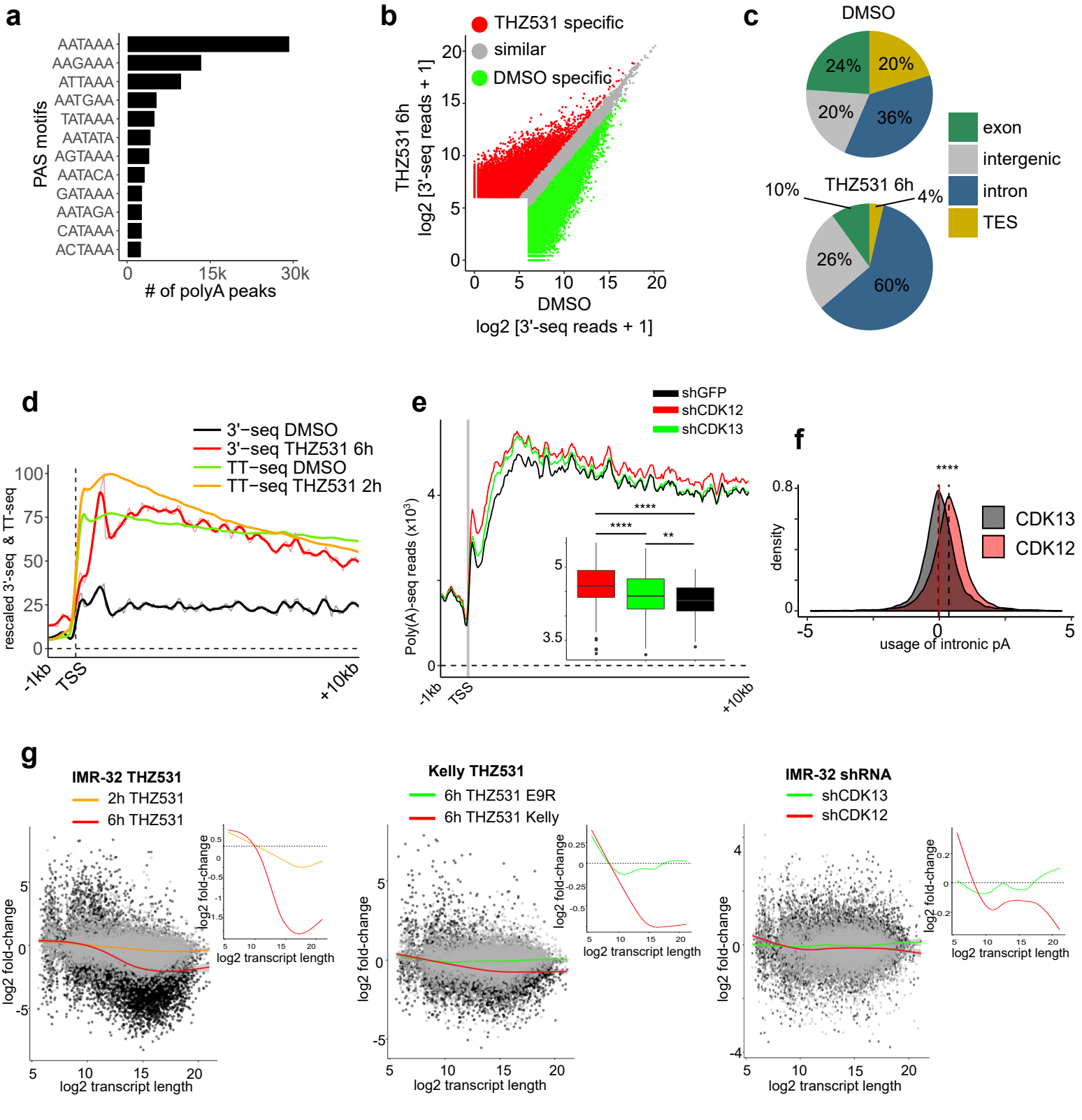
Supplementary Figure 3



Supplementary Figure 4

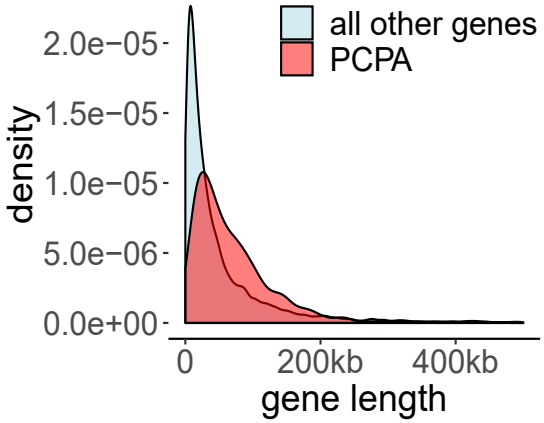


Supplementary Figure 5



Supplementary Figure 6

a

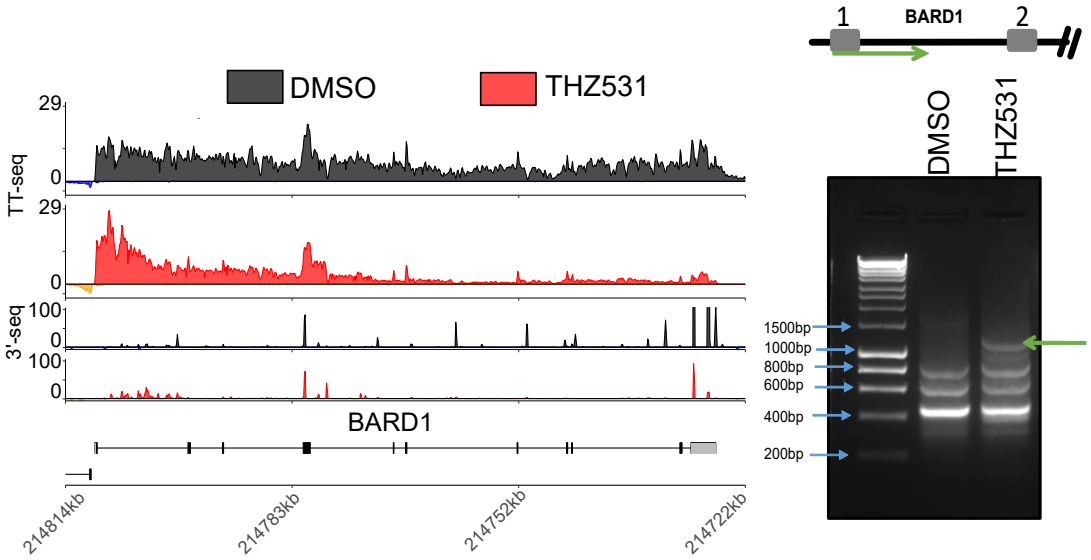


b

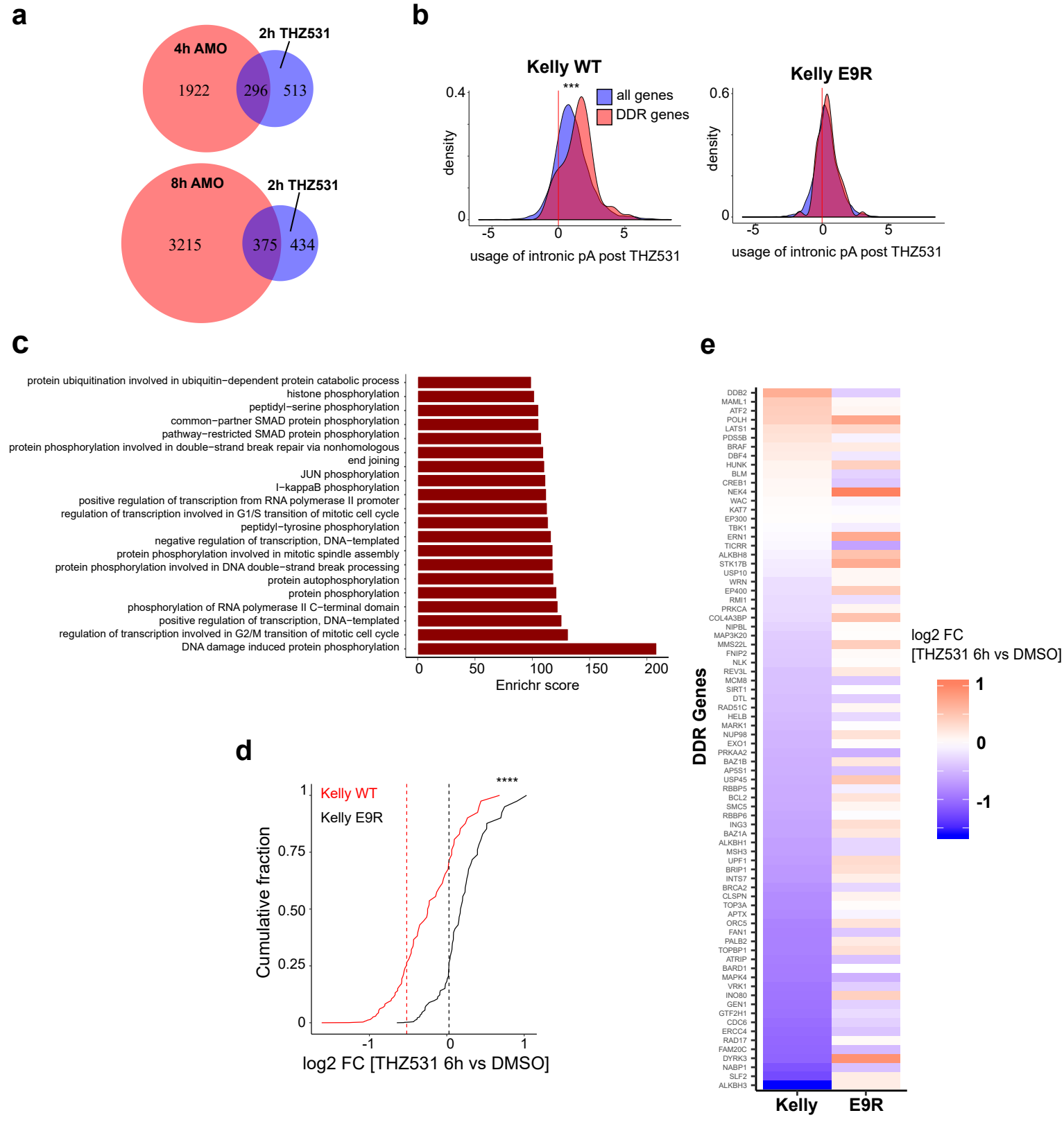
DNA Damage Response genes within the 809 genes that underwent PCPA following THZ531 treatment

short	medium-short	medium-long			long			
NABP1	ERCC4	GEN1	POLH	CLSPN	FAM20C	EP300	VRK1	BRIP1
AP5S1	DYRK3	RBBP6	RAD51C	TBK1	PRKAA2	WAC	MARK1	REV3L
APTX	SLF2	DBF4	TOP3A	USP45	DTL	INTS7	INO80	NIPBL
	CDC6	SIRT1	MAML1	TICRR	CREB1	SMC5	FNIP2	PDS5B
	UPF1	ALKBH1	GTF2H1	LATS1	ERN1	ATF2	WRN	BCL2
	ATRIP	RBBP5	RAD17	NEK4	USP10	BLM	MMS22L	BRAF
	RMI1	STK17B	HELB	ALKBH8	ORC5	NUP98	COL4A3BP	MSH3
	DDB2	PALB2	KAT7	TOPBP1	BAZ1B	BAZ1A	MAP3K20	PRKCA
	ING3	FAN1	EXO1		BARD1	EP400	NLK	
		ALKBH3	MCM8		BRCA2	HUNK	MAPK4	

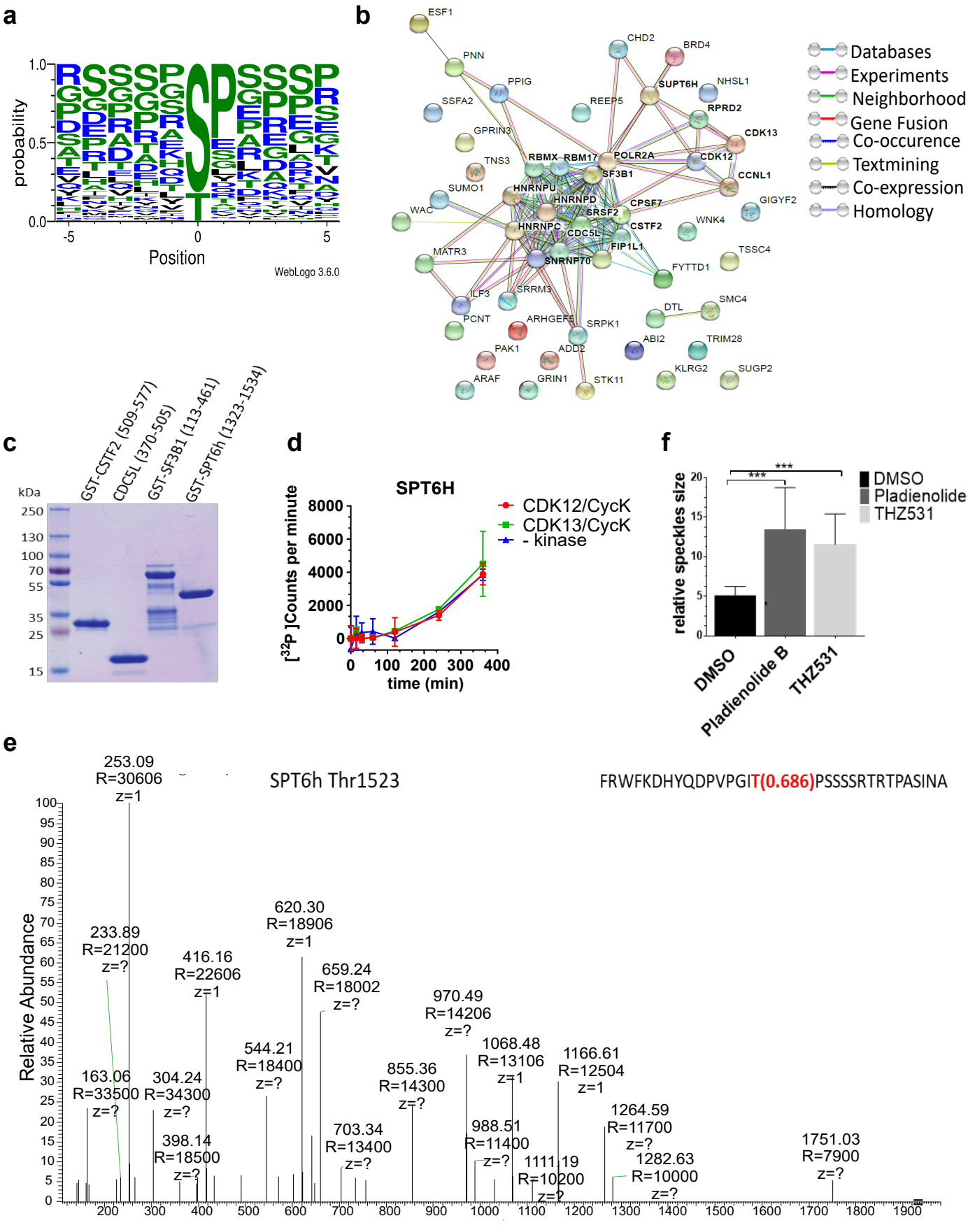
c



Supplementary Figure 7



Supplementary Figure 8



METHODS

Cell Culture. Human neuroblastoma (NB) cells (Kelly, IMR-32, IMR-5, LAN-1, LAN-5, NGP and SK-N-AS) were obtained from the Children's Oncology Group cell line bank and genotyped at the DFCI Core Facility. The cell lines were authenticated through STR analyses. The Kelly E9R NB cell line harbors a single point mutation in CDK12 at the cysteine 1039 covalent binding site of THZ531. Specifically, this mutation was acquired spontaneously in Kelly NB cells upon exposure to escalating doses of CDK12 inhibitor, E9 over the course of few months as previously reported[60]. Human lung (IMR-90) and skin fibroblasts (BJ) were kindly provided by Dr. Richard Gregory (Boston Children's Hospital). NIH3T3 cells were purchased from the American Type Culture Collection (ATCC). NB cells were grown in RPMI (Invitrogen) supplemented with 10% FBS and 1% penicillin/streptomycin (Invitrogen). IMR-90, BJ and NIH3T3 cells were grown in DMEM (Invitrogen) supplemented with 10% FBS and 1% penicillin/streptomycin. All cell lines were routinely tested for mycoplasma.

Compounds. THZ531 was prepared by Dr Nathanael Gray's laboratory. The splicing inhibitor, Pladienolide B was purchased from Santa Cruz Biotechnology.

Cell viability assay. Cells were plated in 96-well plates at a seeding density of 4 x 10^3 cells/well. After 24 h, cells were treated with increasing concentrations of THZ531 (10 nM to 10 μM). DMSO solvent without compound served as a negative control. After 72 h incubation, cells were analyzed for viability using the CellTiter-Glo Luminescent Cell Viability Assay (Promega) according to the manufacturer's instructions. All proliferation assays were performed in biological triplicates and error bars represent mean ± SD. Drug concentrations that inhibited 50% of cell growth (IC_{50}) were determined using a nonlinear regression curve fit using GraphPad Prism 6 software.

Fluorescence-Activated Cell Sorting Analysis (FACS). For cell cycle and DNA damage analysis, cells were treated with DMSO or THZ531, 400nM. After 2, 6 and 24h, cells were trypsinized and fixed in ice-cold 70% ethanol overnight at -20ºC. After washing with ice-cold phosphate-buffered saline (PBS), cells were incubated in PBS-0.5% Tween-20 with γ-H2AX antibody overnight at 4°C. Cells were subsequently washed and incubated with Alexa-488-conjugated secondary antibody for 45 min and then treated with 0.5 mg/ml RNAse A (Sigma-

Aldrich) in combination with 50 µg/ml propidium iodide (PI, BD Biosciences). For apoptosis analysis cells were harvested and stained with PI and FITC-Annexin V according to the manufacturer's protocol (BD Biosciences). All FACS samples were analyzed on a FACS-Calibur (Becton Dickinson) using Cell Quest software (Becton Dickinson). A minimum of 50,000 events was counted per sample and used for further analysis. Data were analyzed using FlowJo software.

shRNA Knockdown. pLKO.1 plasmids containing shRNA sequences targeting CDK12 (sh#1: TRCN0000001795; sh#2 TRCN0000197022), CDK13 (sh#1: TRCN0000000701; sh#2: TRCN0000000704) and GFP were obtained from the RNAi Consortium of the Broad Institute (Broad Institute, Cambridge, MA), knockdowns were performed as described previously[61]. Briefly, the constructs were transfected into HEK293T cells with helper plasmids: pCMV-dR8.91 and pMD2.G-VSV-G for virus production. Cells were then transduced with virus, followed by puromycin selection for two days.

Western Blotting. Cells were collected by trypsinization and lysed at 4°C in NP40 buffer (Invitrogen) supplemented with complete protease inhibitor cocktail (Roche), PhosSTOP phosphatase inhibitor cocktail (Roche) and PMSF (1mM). Protein concentrations were determined with the Biorad DC protein assay kit (Bio-Rad). Whole cell protein lysates were resolved on 4%–12% Bis-Tris gels (Invitrogen) and transferred to nitrocellulose membranes (Bio-Rad). After blocking nonspecific binding sites for 1 h using 5% dry milk (Sigma) in Tris-buffered saline (TBS) supplemented with 0.2% Tween-20 (TBS-T), membranes were incubated overnight with primary antibody at 4ºC. Chemiluminescent detection was performed with the appropriate secondary antibodies and developed using Genemate Blue ultra autoradiography film (VWR).

Immunofluorescence microscopy. Cells were seeded on glass coverslips in six-well plates at a seeding density of 1 x 10^6 cells/well. After 24 h, cells were treated with DMSO or 400nM of THZ531 for 6 or 24 h. Additionally, for the RAD51 staining, cells were irradiated (8 Gy) using a γ-cell 40 irradiator with a cesium source (Best Theratronics, Ltd). Six hours after irradiation cells were washed in PBS and fixed in 3.7% formaldehyde in PBS for 15 min at room temperature (RT). Cells were permeabilized in 0.1% Triton X-100 in PBS for 5 min. Subsequently, cells were extensively washed and incubated with PBS containing 0.05% Tween-20 and 5% BSA

(PBS-Tween-BSA) for 1 h to block nonspecific binding. Cells were then incubated overnight at 4°C with anti-RAD51 or anti-SC-35 (SRSF2) primary antibodies in PBS-Tween-BSA, extensively washed and incubated for 45 min with AlexaFluor 488-conjugated secondary antibodies and counterstained with DAPI. Images were acquired on a Zeiss AXIO Imager Z1 fluorescence microscope using a x63 immersion objective, equipped with AxioVision software. Nuclei with > 5 RAD51 foci were considered positive and 100 nuclei per condition were analyzed.

Target Engagement Assay. Cells were treated with THZ531 or DMSO for 6 h at the indicated doses. Subsequently, total cell lysates were prepared as for western blotting. To IP CDK12 and CDK13, 1 mg and 4 mg respectively of total protein was incubated with 1μM of biotin-THZ531 at 4°C overnight. Subsequently, lysates were incubated with streptavidin agarose (30 μl) for 2 h at 4°C. Agarose beads were washed 3x with cell lysis buffer and boiled for 10 min in 2x gel loading buffer. Proteins were resolved by WB. 50 μg of total protein was used as a loading control.

Stable Isotope Labeling by Amino Acids in Cell Culture (SILAC). IMR-32 and Kelly cells were grown in arginine- and lysine-free RPMI with 10% dialyzed FBS supplemented with either [$^{13}C_6$,$^{15}N_2$] lysine (100 mg/liter) or [$^{13}C_6$,$^{15}N_4$] arginine (100 mg/liter) (Cambridge Isotope Laboratories, Inc.) (heavy population) or identical concentrations of isotopically normal lysine and arginine (light population) for at least six cell doublings. Heavy-labeled cells were incubated in THZ531 (400nM) for 2 h and light-labeled cells were incubated in DMSO solvent as a control. After inhibitor treatment, cells were collected by trypsinization and counted. Equal numbers of heavy and light cells were mixed, washed twice in PBS, snap-frozen, and stored at −80°C until lysis.

Phosphopeptide purification. Phosphopeptide enrichment was performed using titanium dioxide microspheres as previously described[62]. Briefly, lyophilized peptides were dissolved in 50% acetonitrile (ACN; Honeywell)/2M lactic acid (Lee Biosolutions), incubated with 1.25 mg TiO2 microspheres (GL Sciences) per 1 mg peptide digest and vortexed at 75% power for 1 h. Microspheres were washed twice with 50% ACN/2M lactic acid and twice with 50% ACN/0.1% TFA. Phosphopeptides were eluted with 50 mM K_2HPO_4 (Sigma) pH 10 (adjusted with

ammonium hydroxide; Sigma). Formic acid (EMD) was added to the eluates to a concentration of 1.7%. The acidified phosphopeptides were desalted using a C18 solid-phase extraction (SPE) cartridge and the eluate was vacuum centrifuged to dryness.

Offline HPLC pre-fractionation. Approximately 120 µg phosphopeptides were resuspended in 0.1% TFA (Trifluoroacetic acid) and fractionated via pentafluorophenyl chromatography as previously described[63]. The 48 collected fractions were reduced to 16 by combining every 16th fraction, vacuum centrifuged to dryness and stored at -80°C prior to analysis by LC-MS/MS.

LC-MS/MS analysis. LC-MS/MS analysis was performed on an Orbitrap Fusion Tribrid mass spectrometer (ThermoFisher Scientific, San Jose, CA) equipped with an EASY-nLC 1000 ultra-high pressure liquid chromatograph (ThermoFisher Scientific, Waltham, MA). Phosphopeptides were dissolved in loading buffer (5% methanol (Fisher)/1.5 % formic acid) and injected directly onto an in-house pulled polymer coated fritless fused silica analytical resolving column (40 cm length, 100µm inner diameter; PolyMicro) packed with ReproSil, C18 AQ 1.9 µm 120 Å pore (Dr. Maisch). Phosphopeptides in 3 µl loading buffer were loaded at 650 bar pressure by chasing onto the column with 10µl loading buffer. Samples were separated with a 90-min. gradient of 4 to 33% LC-MS buffer B (LC-MS buffer A: 0.125% formic acid, 3% ACN; LC-MS buffer B: 0.125% formic acid, 95% ACN) at a flow rate of 330 nl/min. The Orbitrap Fusion was operated with an Orbitrap MS1 scan at 120K resolution and an AGC target value of 500K. The maximum injection time was 100 milliseconds, the scan range was 350 to 1500 m/z and the dynamic exclusion window was 15 seconds (±15 ppm from precursor ion m/z). Precursor ions were selected for MS2 using quadrupole isolation (0.7 m/z isolation width) in a “top speed” (2 second duty cycle), data-dependent manner. MS2 scans were generated through higher energy collision-induced dissociation (HCD) fragmentation (29% HCD energy) and Orbitrap analysis at 15K resolution. Ion charge states of +2 through +4 were selected for HCD MS2. The MS2 scan maximum injection time was 60 milliseconds and AGC target value was 60K.

Peptide spectral matching and bioinformatics. Raw data were searched using COMET[64] against a target-decoy version of the human (*Homo sapiens*) proteome sequence database (UniProt; downloaded 2013; 20,241 total proteins) with a precursor mass tolerance of ± 1.00 Da and requiring fully tryptic peptides with up to 3 missed cleavages, carbamidomethyl cysteine as a fixed modification and oxidized methionine as a variable modification. For SILAC experiments, the additional masses of lysine and arginine isotope labels were searched as variable modifications. Phosphorylation of serine, threonine and tyrosine were searched with up to 3 variable modifications per peptide, and were localized using the phosphoRS algorithm[65]. The resulting peptide spectral matches were filtered to <1% false discovery rate (FDR) by defining thresholds of decoy hit frequencies at particular mass measurement accuracy (measured in parts per million from theoretical), XCorr and delta-XCorr (dCn) values.

Antibodies. The following antibodies were used: RNAPII CTD S2 (Bethyl cat# A300-654A); RNAPII CTD S5 (Bethyl cat# A300-655A); RNAPII (Santa Cruz cat# sc-899); RNAPII CTD S7 (Millipore cat#041570) RNAPII CTD Thr4 (Active Motif cat# 61361) cleaved PARP (Cell Signaling cat# 9541); GAPDH (Cell Signaling cat# 2118S); CDK12 (Cell Signaling cat# 11973S); CDK13 (Bethyl cat# A301-458A); γ-H2AX (Cell Signaling cat# 9718), RAD51 (GeneText cat# GTX70230), BRCA1 (Cell signaling cat# 9010S), BARD1 (Santa Cruz Biotechnology cat#sc11438), Alexa-488 (Molecular Probes cat#A11008), SC-35 (Abcam cat# ab11826).

RT-PCR. Total RNA was isolated with the RNAeasy Mini kit (QIAGEN). One μg of purified RNA was reverse transcribed using Superscript III First-Strand (Invitrogen) with random hexamer primers following the manufacturer's protocol. Quantitative PCR was carried out using the QuantiFast SYBR Green PCR kit (Qiagen) and analyzed on an Applied Biosystems StepOne Real-Time PCR System (Life Technologies). Each individual biological sample was qPCR-amplified in technical triplicate and normalized to GAPDH as an internal control. Relative quantification was calculated according to the ΔΔCt relative quantification method. Error bars indicate ± SD of three replicates. Primers sequences are available on request.

RNA extraction and synthetic RNA spike-in for gene expression analysis. Cells were treated with 400 nM of THZ531 or with DMSO for 6 h. Cell numbers were determined prior to lysis and RNA extraction. Biological

duplicates (5 million cells per replicate) were collected and homogenized in 1 ml of TRIzol Reagent (Invitrogen) and purified using the mirVANA miRNA isolation kit (Ambion) following the manufacturer's instructions. Total RNA was treated with DNA-free™ DNase I (Ambion), spiked-in with ERCC RNA Spike-In Mix (Ambion), and analyzed on an Agilent 2100 Bioanalyzer (Agilent Technologies) for integrity. RNA was hybridized to Affymetrix GeneChip_ PrimeView Human Gene Expression arrays (Affymetrix).

Transient Transcriptome Sequencing. Cells were treated with DMSO or 400 nM of THZ531 for 30 min and 2 h. Cells were labeled in media for 10 min with 500 μM 4-thiouridine (4sU, Sigma-Aldrich). RNA extraction was performed with TRIzol (Ambion) following the manufacturers' instructions. Total RNA was treated with DNase I (Invitrogen). Subsequently, the purified RNA was fragmented on a BioRuptor Next Gen (Diagenode) at high power for one cycle of 30''/30'' ON/OFF. Fragmented samples were subjected to labeled RNA purification as previously described[66]. Labeled fragmented RNA was spiked-in with ERCC RNA Spike-In Mix (Ambion) and analyzed on an Agilent 2100 Bioanalyzer (Agilent Technologies) for integrity. Sequencing libraries were prepared with the RNA-seq library kit (TruSeq Stranded Total RNA RiboZero Gold, Illumina) as per the manufacturers' instructions. All samples were sequenced on a HiSeq 2500 sequencer.

Poly(A) 3′-end sequencing. Cells were treated with DMSO or THZ531 (400 nM in IMR-32 cells; 200nM in Kelly and Kelly E9R cells) for 2 and 6 h. RNA extraction was performed with TRIzol (Ambion) following the manufacturers' instructions. Total RNA was treated with DNase I (Invitrogen). Sequencing libraries were prepared with the RNA-seq library kit (QuantSeq 3' mRNA Sequencing REV, Lexogen) following the manufacturers' instructions. All samples were sequenced on a HiSeq 2500 sequencer.

In vitro kinase assay. Recombinant CDK12/CycK complex was prepared from baculovirus infected Sf9 cells as described[67]. Substrate proteins CstF64 (aa 509-577), CDC5L (aa 370-505), SPT6H (aa 1434-1544) and SF3B1 (aa113-462) were expressed as GST-fusion proteins in E. coli and purified to homogeneity. Radioactive kinase reactions were performed with 0.2 μM CDK12/CycK or CDK13/CycK and 100 μM each of substrate protein, and 1 mM ATP at 30°C for 30 min in kinase buffer as described[67]. Reactions were spotted onto P81 Whatman

paper squares, washed three times and radioactivity counted on a Beckman Scintillation Counter (Beckman-Coulter) for 1 min. Measurements were performed in triplicate and are represented as mean ± S.D.

Peptide mass fingerprinting. GST-tagged proteins were resolved on a 12% SDS-PAGE gel and stained with Coomassie brilliant blue. Protein bands were cut from the SDS-PAGE gel and submitted for mass spectrometry analysis to the Bioanalytical Mass Spectrometry Group, Max Planck Institute for Biophysical Chemistry, Gottingen**,** Germany.

3′ RACE PCR. Cells were treated with 400nM of THZ531 or with DMSO for 6 h. RNA extraction was performed with TRIzol (Ambion) following the manufacturers' instructions. Total RNA was treated with DNase I (Invitrogen). Subsequently, total RNA was poly(A) selected using oligo-dT dynbeads (Invitrogen). RNA was reverse transcribed using 3′-RACE adaptor oligonucleotide (FirstChoice RLM-RACE Kit; Life Technologies). Nested PCR was performed using Phusion High Fidelity DNA Polymerase (New England Biolabs). PCR products were resolved on a 1.3% agarose gel, purified using a gel extraction kit (Qiagen), and sequenced. PCR primer sequences available upon request.

Gene expression analysis. Microarray data were analyzed using a custom CDF file (GPL16043) that contained the mapping information of the ERCC probes used in the spike-in RNAs. The arrays were normalized according to previously described protocols[68]. Briefly, all chip data were imported in R (version 3.0.1) using the affy package[69], converted into expression values using the expresso command, normalized to take into account the different numbers of cells and spike-ins used in the different experiments and renormalized using loess regression fitted to the spike-in probes. Sets of differentially expressed genes were obtained using the limma package[70] and a False Discovery Rate (FDR) of 0.05. Statistical comparisons of distributions of fold changes were done using the Mann-Whitney U test.

TT-sequencing data processing. For each sample, paired-end 75 bp reads were obtained and mapped to the human genome (GRCh38) and ERCC spike-in sequences. An average of 50 M (~90%) read pairs were mapped with STAR (version STAR_2.5.1b_modified) with default parameters. Only high quality and properly paired

reads were retained for further analysis using samtools (v1.3.1) with parameters "*-q 7* and *–f 83,99,147,163"*. To normalize for library size and sequencing depth variation, individual spike-in reads were counted with *samtools idxstats*. This was used as input to calculate a sample-specific size factor with *estimateSizeFactorsForMatrix* (DESeq2). To create strand-specific sample coverage profiles in 100 bp bins, we used bamCoverage (DeepTools v2.5.4) with previously calculated size factors and parameters "*--scaleFactor --normalizeUsingRPKM –filterRNAstrand –bs 100"*. Genome-wide correlation of biological replicates was calculated using Spearman's rank coefficient and visualized using scatterplots and heatmaps. These results showed high reproducibility for each condition and hence, for all analyses except differential expression and transcript usage, replicates were merged using *samtools merge* and processed again as described for the individual replicates.

Poly(A) 3′-sequencing data processing. For each sample single-end 100 bp reads were obtained and filtered using bbduk.sh from BBMap (v37.00) and parameters *"k=13 ktrim=r useshortkmers=t mink=5 qtrim=r trimq=20 minlength=75 ref=truseq_rna.fa.gz"* to remove potential adaptor contamination or low-quality reads. High-quality reads were subsequently mapped to the human genome (GRCh38) with STAR (version STAR_2.5.1b_modified) and the following parameters *"--outFilterType BySJout --outFilterMultimapNmax 20 --alignSJoverhangMin 8 --alignSJDBoverhangMin 1 --outFilterMismatchNmax 999 --outFilterMismatchNoverLmax 0.1 --alignIntronMin 20 --alignIntronMax 1000000 --alignMatesGapMax 1000000 --outSAMattributes NH HI NM MD --outSAMtype BAM SortedByCoordinate"*. To create strand-specific sample coverage profiles in 50bp bins, we used bamCoverage (DeepTools v2.5.4) with parameters "*--normalizeUsingRPKM –filterRNAstrand –bs 50"*. Genome-wide correlation of biological replicates was calculated using Spearman's rank coefficient and visualized using scatterplots and heatmaps. These results showed high reproducibility for each condition and hence, for all visualizations, replicates were merged using *samtools merge* and processed again as described for the individual replicates. Dataset-specific steps: (i) for the Kelly WT and Kelly E9R samples ERCC spike-in sequences were added and included in the downstream analysis to allow the detection of absolute gene expression level differences in another NB cell line. (ii) The CDK12/13 shRNA-mediated knockdown samples were processed in two different batches which displayed variable

outcomes in sequence quality and depth, hence, to adjust for technical confounding effects, the samples were first randomly downsampled to the lowest individual library size.

Poly(A) 3′-sequencing peak identification and filtering. Poly(A)-seq peaks were called using MACS2 (v 2.1.1) and parameters *"--nomodel --extsize 100 --shift 0"* using a merged bam file of all samples. To identify false positive poly(A) peaks, two criteria were used: 1) the presence of the potential PAS motifs (AATAAA, ATTAAA, AGTAAA, TATAAA, AATATA, AATACA, CATAAA, GATAAA, AATGAA, ACTAAA, AAGAAA, AATAGA) computed in a 100 bp window upstream of the peak in a strand-specific manner, and 2) the presence of a genomic 25-adenine (A) stretch with a maximum of 3 mismatches computed in a 50 bp window downstream of the peak in a strand-specific manner. Peaks were removed if they were not associated with a PAS motif but were associated with a genomic stretch of A's. Reads associated with these peaks were subsequently removed from the original mapped reads with the command "*samtools –L regions_to_remove.bed –U output.bam"*. Strand specific coverage files were recomputed as described before. Retained Poly(A)-seq peaks were annotated in a step-wise manner; first, peaks were considered to be associated with the 3' UTR if they were within the vicinity of the transcription end site (TES, -200 bp to + 600 bp), next, the remaining peaks were considered to be intergenic or genic and, in the latter case, overlapping with an exon or intron. If a peak overlapped multiple transcripts, priority was given to protein-coding transcripts followed by longer transcripts. For metagene plots genes were represented by the isoform that showed the highest combined 3'-UTR expression level.

Transcript selection and custom genome annotation. Kallisto (v0.43.1) with parameters "*--bootstrap-samples --rf-stranded"* was used to determine the relative expression levels of all annotated transcripts as transcripts per million (TPM) (GencodeV27, GRCh38.p10). To reduce noise for downstream analyses, low-expressed genes (gene TPM < 2) or infrequently used transcripts (fraction of transcript < 0.2, except if transcript TPM > 5) were removed. A custom genome annotation was created by only retaining the detected transcripts. For each gene, all individual transcripts were merged using the reduce function of the GenomicRanges package in R to create a reduced exonic or intronic representation.

Differential transcript usage. Differential transcript usage between DMSO and THZ531-treated samples at 2h was determined with the rats package in R using count estimates from Kallisto and further filtered based on our custom gene annotation.

Alternative splicing. To extract alternative splicing events the TT-seq paired-end reads were first re-mapped with STAR as described before, except soft-clipping was excluded by setting the parameter *"--alignEndsType"* to *"EndToEnd"* to favor reads spanning the exon-intron border. Next, we used the rMATS tool (v4.0.1) with the default settings to identify statistically significant alternative events (FDR < 0.05).

Intron retention index. The intron retention (IR) index is the log2 ratio of the exon and intron ratios calculated on the TT-seq normalized coverage in THZ531- and DMSO-treated samples. Only genes with a minimum of 10 exonic and 5 intronic TT-seq reads in either THZ531 or DMSO treated cells were included for this analysis. In short:

- Exon ratio = (exon coverage THZ531 +1) / (exon coverage DMSO + 1)

- Intron ratio = (intron coverage THZ531 + 1) / (intron coverage DMSO + 1)

- IR index = log2 (Intron ratio / Exon ratio)

The Fisher's exact test was used to determine significant intron loss (adjusted *p*-value < 0.05 and IR index < -1) or retention (adjusted *p*-value < 0.05 and IR index > 1).

Sample-specific poly(A) 3′-seq peaks and genomic distribution. To calculate sample-specific poly(A) 3′-seq peaks, only peaks with a $-\log_{10}$ q-score ≥ 5 were retained. For each peak, overlapping reads were counted for each condition and a log-ratio score was calculated as log2 (THZ531_6h+1/DMSO+1). Peaks with a minimum number of 64 reads and a log-ratio score > 1 or < -1 were considered THZ531_6h- and DMSO-specific respectively. Each peak was assigned to only one genomic region based on overlap with our custom gene annotation in a ranked order, i.e. annotated TES, exon, intron or intergenic.

Differential expression. Pairwise differential expression for exonic regions between DMSO- and THZ531-treated samples was calculated in the following manner: a 5′ upstream (-50 bp to -2050 bp of TSS) and 3′ downstream (+50 bp to +2050 bp of TES) 2 kb window was created and regions overlapping with genes on the same strand were removed. Only regions with a final minimum length of 200 bp were retained for further analysis. Genomic locations for exonic regions were converted to an saf format to calculate gene counts for each region using FeatureCounts (v1.5.0-p1). To detect differentially expressed genes for each genomic region the DESeq2 package in R was used with the size factors calculated previously (see TT-seq data processing). A gene with an absolute log2 fold-change > 1 and an adjusted p-value < 0.1 was considered significant.

Differential global expression change of aggregated DDR gene set. A combined set of genes that are part of the DNA damage response (DDR) was created by aggregating genes assigned to any DDR pathway in the databases found at https://www.mdanderson.org/documents/Labs/Wood-Laboratory/human-dna-repair-genes.html and http://repairtoire.genesilico.pl (see Table 3). To identify if these genes as a whole were more downregulated, 1000 random and equal-sized gene sets were generated and a distribution of the average level of expression change was plotted and compared to that of the initial DDR gene set to calculate a z-score.

Gene biotype and size selection. Gene biotypes assigned by Gencode were simplified in a 2-step manner. First, only gene biotypes with a minimum of 20 members were considered. Next gene lengths of all detected non-coding genes (i.e. excluding protein-coding) were clustered using kmeans in 2 groups (short vs. long non-coding genes). Together, this resulted in 3 groups selected on biotype and gene length: 1) protein-coding genes 2) long non-coding genes (lincRNA, antisense_RNA, processed_transcript, sense_intronic, transcribed_unitary_pseudogene, TEC (to be experimentally confirmed), transcribed_processed_pseudogene, transcribed_unprocessed_pseudogene, unprocessed_pseudogene), and 3) short non-coding genes (snRNA, scaRNA, snoRNA, Mt_tRNA, misc_RNA, processed_pseudogene, rRNA). Protein-coding genes were further stratified into 4 length classes based on the quartiles of length distribution, i.e. long (>64.5 kb), medium-long (26.4 – 64.5 kb), medium-short (9.9 – 26.4 kb) and short (< 9.9 kb). The latter group was further divided into 3 equal groups based on the 0.33, 0.66 and 1 quantiles of only the short gene length distribution, resulting in short

(9.9 – 6.4 kb), very short (6.4 – 3.4 kb) and ultra short (< 3.4 kb) groups. Simple linear regression and Spearman correlation coefficient between log2 scaled length and log2 fold-change of exonic reads for genes was performed in R.

Metagene profiles. A gene metaprofile was created by dividing each gene (from TSS to TES) into 50 equally sized bins; 2 kb upstream and downstream flanking regions were binned in bins of 100 bp. Bedgraph files with normalized reads from TT-seq or poly(A) 3′-seq were used to calculate read density (RPM/bp) across those bins and subsequently summarized for all genes. To create a TSS or TES metaprofile, we followed an analogous approach with variable upstream and downstream flanking regions and summarized bins of 50 bp. To compare TT-seq and poly(A) 3′-seq profiles, calculated read densities were rescaled between 1-100.

Inference of proximal RNA polymerase dynamics. Transcription dynamics were calculated in 50 bp bins in the 500 bp upstream and 10 kb downstream flanking regions of the TSS. Change in read accumulation was calculated by normalized TT-seq read subtraction [THZ531 – DMSO] and the first derivative was computed to obtain the rate of accumulation change. Both the change and the rate of change were rescaled between 0 and 1 and a loess smoothing curve was then fitted for visualization purposes.

Correlation of transcript length and 3' expression changes. To identify differential expressed genes based on 3' poly(A)-sequencing, all counts for 3' UTR-associated polyadenylation sites were summarized per gene. This data matrix was log2 normalized and used to identify differential expression and fold-changes with the limma package in R. Correlation between fold-changes and transcript length was performed on the highest expressed transcript for each gene in the control condition. A generalized additive model (GAM) smoothing curve was fitted to each treatment to observe global changes and for visualization purposes.

PCPA analysis. Treatment-induced PCPA for protein-coding genes was calculated as in Oh et al[71] with minor modifications. To determine whether a gene exhibits a coverage profile expected with PCPA, two scores were calculated. First, for each gene, we calculated an exon-score to determine if there was an increased loss of reads at the last exon compared to the first exon with THZ531 treatment: last exon [log2 (THZ531_2h / DMSO)] – first

exon [log2 (THZ531_2h / DMSO)]. Next, an iQ-score to determine if there was an increased number of reads in the first quarter (iQ1) of the region between a gene's first 5′ splice site and the last 3′ splice site and the last quarter (iQ4) was calculated for each gene in the THZ531-treated samples: log2(iQ1 / iQ4). Genes were considered to undergo THZ531-induced PCPA with an exon-score < -1 and an iQ-score > 1.

Intronic polyadenylation usage. For each transcript (TPM > 1) the reads of all intronic and 3′ UTR-associated poly(A) sites were summarized. To compare the change and usage of intronic versus 3′ UTR-associated poly(A) sites between different treatments, an odds ratio (OR) was calculated for each treatment sample but excluding transcripts that had no intronic poly(A) sites in either treatment. A two-sample Kolmogorov-Smirnov test was then used to detect changes in OR distributions between different treatments.

Correlation of transcript length and number of intronic polyadenylation sites. To identify the relationship between the number of identified polyadenylation sites and transcript length, a polynomial regression curve (y ~ poly(x,2)] was fitted for all genes or DDR genes only. A Wilcoxon Rank Sum test was used to determine if the difference between predicted values for DDR genes between the two models [prediction DDR– prediction all genes] was significantly different.

Genetic determinants analysis. Known U1 (GGTGAG, GGTAAG and GTGAGT), PAS (AATAAA) motifs and GC content percentages were computed for each gene along the entire gene axis (TSS to TES) using the Biostrings package in R. A Wilcoxon Rank Sum test and Cohen's *d* effect size were used to determine individual differences between all genes and genes with PCPA for each selected genetic determinant (gene length, length of first intron, number of introns, GC content, expression and ratio between U1 and PAS).

Splice site conservation analysis. Calculation of splice site conservation scores was performed as previously described[72] with modifications. In brief, position weight matrices (PWMs) for 5′ and 3′ splice sites were created using all introns that contain the established 5′ GT and 3′ AG sequence. For the 5′ and 3′ splice sites (ss) 9 (-2 bp:6 bp of 5′ss) and 15 bp (-14 bp of 3′ss) respectively were used. Next, introns with and without intronic polyadenylation sites (intronic poly(A) > 0) were scored for both the 5′ and 3′ PWM for their respective splice

sites. A combined score for each intron was computed by summarizing the scores of the 5′ and 3′splice site. The Wilcoxon Rank Sum test and Cohen's *d* effect size were used to determine biologically meaningful differences between introns with and without an intronic polyadenylation.

Enrichment analysis. Gene ontology enrichment for selected gene sets was performed using the enrichR package in R. The Enrichr score[73] is the combined score of the adjusted *p*-value and the z-score using the Fisher's exact test. Enrichment of individual gene sets was considered significant if the adjusted *p*-value < 0.01, unless stated otherwise. The Fisher's exact test was used to determine significant overlap between other publicly available datasets.

Genomic visualization. To visualize coverage tracks a custom build visualization tool was used (github.com/RubD/GeTrackViz2).

Data availability. Microarray, TT-seq, Poly(A) 3′-seq datasets have been deposited in the Gene Expression Omnibus (GEO), accession number GSE113314. The SILAC dataset has been deposited in the ProteomeXchange Consortium, accession number PXD009533. All other data are available from the corresponding author upon request.

References

60 Gao, Y. *et al.* Overcoming Resistance to the THZ Series of Covalent Transcriptional CDK Inhibitors. *Cell Chem Biol* **25**, 135-142 e135, doi:10.1016/j.chembiol.2017.11.007 (2018).

61 Chipumuro, E. *et al.* CDK7 inhibition suppresses super-enhancer-linked oncogenic transcription in MYCN-driven cancer. *Cell* **159**, 1126-1139, doi:10.1016/j.cell.2014.10.024 (2014).

62 Kettenbach, A. N. & Gerber, S. A. Rapid and reproducible single-stage phosphopeptide enrichment of complex peptide mixtures: application to general and phosphotyrosine-specific phosphoproteomics experiments. *Anal Chem* **83**, 7635-7644, doi:10.1021/ac201894j (2011).

63 Grassetti, A. V., Hards, R. & Gerber, S. A. Offline pentafluorophenyl (PFP)-RP prefractionation as an alternative to high-pH RP for comprehensive LC-MS/MS proteomics and phosphoproteomics. *Anal Bioanal Chem* **409**, 4615-4625, doi:10.1007/s00216-017-0407-6 (2017).

64 Eng, J. K., Jahan, T. A. & Hoopmann, M. R. Comet: an open-source MS/MS sequence database search tool. *Proteomics* **13**, 22-24, doi:10.1002/pmic.201200439 (2013).

65 Taus, T. *et al.* Universal and confident phosphorylation site localization using phosphoRS. *J Proteome Res* **10**, 5354-5362, doi:10.1021/pr200611n (2011).

66 Dolken, L. *et al.* High-resolution gene expression profiling for simultaneous kinetic parameter analysis of RNA synthesis and decay. *RNA* **14**, 1959-1972, doi:10.1261/rna.1136108 (2008).

67 Bosken, C. A. *et al.* The structure and substrate specificity of human Cdk12/Cyclin K. *Nat Commun* **5**, 3505, doi:10.1038/ncomms4505 (2014).

68 Loven, J. *et al.* Revisiting global gene expression analysis. *Cell* **151**, 476-482, doi:10.1016/j.cell.2012.10.012 (2012).

69 Gautier, L., Cope, L., Bolstad, B. M. & Irizarry, R. A. affy--analysis of Affymetrix GeneChip data at the probe level. *Bioinformatics* **20**, 307-315, doi:10.1093/bioinformatics/btg405 (2004).

70 Smyth, G. K., Yang, Y. H. & Speed, T. Statistical issues in cDNA microarray data analysis. *Methods Mol Biol* **224**, 111-136, doi:10.1385/1-59259-364-X:111 (2003).

71 Oh, J. M. *et al.* U1 snRNP telescripting regulates a size-function-stratified human genome. *Nat Struct Mol Biol* **24**, 993-999, doi:10.1038/nsmb.3473 (2017).

72 Tian, B., Pan, Z. & Lee, J. Y. Widespread mRNA polyadenylation events in introns indicate dynamic interplay between polyadenylation and splicing. *Genome Res* **17**, 156-165, doi:10.1101/gr.5532707 (2007).

73 Kuleshov, M. V. *et al.* Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res* **44**, W90-97, doi:10.1093/nar/gkw377 (2016).