Ku limits aberrant mRNA splicing promoted by intronic antisense Alu elements.

Giovanni Pascarella[1], Mariia Mikhova[2,3], Gargi Parkhi[4,5], Jared Godfrey[4], Joshua Heyza[3], Tomáš Janovič[3,8], Andrew Olive[4], Piero Carninci[1,6], Jens C Schmidt[3,7,9], and Katheryn Meek[4,5,9].

[1]RIKEN Center for Integrative Medical Sciences (IMS), Yokohama Japan
[2]Department of Biochemistry and Molecular Biology, Michigan State University, East Lansing, MI, U.S.A.
[3]Institute for Quantitative Health Sciences and Engineering, Michigan State University, East Lansing, MI, U.S.A.
[4]Department of Microbiology, Genetics, & Immunology, Michigan State University, East Lansing, MI 48824, U.S.A.
[5]Department of Pathobiology & Diagnostic Investigation, Michigan State University, East Lansing, MI 48824, USA
[6] Human Technopole, Milan, Italy
[7]Department of Obstetrics, Gynecology, and Reproductive Biology, Michigan State University, East Lansing MI, U.S.A.
[8]Current address: Functional Genomics and Proteomics, National Centre for Biomolecular Research, Faculty of Science, Masaryk University, Brno 625 00, Czech Republic.

[9] Correspondence kmeek@msu.edu, schmi706@msu.edu.

Summary

Alu elements are short repeats that occupy approximately 10% of the human genome [1,2]. Saturation of primate genomes with Alu sequences occurred at the prosimian/new-world monkey evolutionary juncture. Alu elements have clearly driven unique aspects of higher primate evolution, but their presence can be detrimental to genomic stability [3]. The expansion of Alu sequences in the genomes of higher primates precisely coincides with a substantial increase in the ubiquitous expression of the three polypeptides of the DNA-dependent protein kinase (DNA-PK), the Ku70/80 heterodimer and DNA-PKcs [4]. Previous work suggests that the elevated levels of Ku70/80 are required to prevent the activation of innate immune signaling pathways triggered by RNA molecules derived from Alu elements [5]. Here we demonstrate that Ku ablation dramatically alters mRNA splicing, by allowing the use of alternative splice sites contained in intronic antisense Alu elements, which are known to directly associate with Ku70/80 [5]. Dysregulation of mRNA splicing precedes cell death and preferentially impacts genes involved in essential RNA metabolism processes, including splicing and ribosome biogenesis, likely impacting cell viability. In addition, we demonstrate that cell death after Ku70 depletion cannot be rescued by expression of its prosimian homologue, which suggests that primate Ku70 has evolved specific molecular features to suppress deleterious effects of an Alu element rich genome. We propose a model in which Ku binding of antisense Alu elements in introns of nascent RNAs modulates the use of alternative splice sites to balance beneficial and detrimental contributions of Alu repeats within primate genomes.

Main

The primary function of the DNA-dependent protein kinase complex is recognition of DNA double-strand breaks (DSBR) during non-homologous end-joining (NHEJ). This complex contains the DNA end-binding heterodimer, Ku70 and Ku80, and the large catalytic subunit, DNA-PKcs [reviewed in [6]]. In NHEJ, DNA-PK initially encircles DNA double-stranded breaks (DSBs) and then facilitates the assembly of NHEJ's ligase complex (Ligase 4, XRCC4, XLF, and PAXX) into several distinct complexes that facilitate end-processing and eventual end ligation [7-9]. It has been well-established that human cells express dramatically higher levels of all three DNA-PK components than non-primate cells [10,11]. We previously reported that the high expression levels of Ku70, Ku80, and DNA-PKcs occurs evolutionarily when new world monkeys diverged from prosimians (primitive primates) [4]. Remarkably this coincides precisely with the expansion of repetitive Alu elements in the genomes of higher primates [4]. Alu elements mediate the vast majority of non-allelic homologous recombination (NAHR) in human cells [12], likely an important factoid that explains many aspects of primate evolution. But NAHR is often deleterious [13,14], and our study suggested that DNA-PK may modestly restrain Alu-mediated NAHR [4].

It has been known for over a decade that Ku80 is essential in human cells, but not rodent cells [15,16]. Although DNA-PKcs is not essential in human cell strains, the lack of DNA-PKcs null patients suggests that DNA-PKcs is essential for organismal viability in humans [4,17]. Recently, new functions, potentially independent of DNA end-joining have been ascribed to the Ku heterodimer [5,18], and it is likely those functions, not its role in NHEJ, that are essential for human cell viability. Hendrickson and colleagues were the first to establish that Ku80 is essential in cultured human cells [15,16]; cell death was attributed to telomeric loss by a hyper-recombination mechanism [19]. The telomere-dependent essentialness of Ku has recently been revisited by Kelly et al. [18], and Zhu et al. [5]. Both studies concluded that telomeric loss did not accurately explain the rapid cell death observed after Ku ablation in human cells. In Kelly et al., the authors could not define a single pathway that was impacted by Ku disruption but proposed that numerous cellular processes were disrupted including RNA related processes, translation, ribosome biogenesis, cell cycle and mitotic regulation. In contrast, Zhu et al. proposed that high levels of Ku are required to bind abundant Alu-encoded RNAs expressed in higher primates, preventing these RNAs from activating innate immune pathways. They suggested that the timely correlation between high levels of DNA-PK and the genomic expansion of Alu elements [4], was an adaptation to restrain Alu-encoded RNAs from activating innate immunity.

Here, we do not observe activation of innate immune signaling pathways in cells depleted of Ku70. In contrast, we find that depletion of Ku70 results in substantial dysregulation of mRNA splicing. The splicing alterations are conserved across multiple cell lines and impact a core set of genes involved in RNA processing, which likely impacts processes essential for cell survival including ribosome biogenesis and mRNA maturation. We propose a unified model in which high levels of Ku70/80 are required to bind antisense Alu elements within nascent RNA introns to prevent the use of cryptic splice sites contained in the antisense Alu sequences.

Results

Cellular ablation of Ku70, or Ku80 but not XLF or DNA-PKcs results in rapid cell death in U2OS cells. Because there is still much to understand about the precise essential requirement for Ku and DNA-PKcs in higher primates, we took advantage of our recently developed reagents including a panel of U2OS cell lines that were gene edited to incorporate a Flag-Halo (Halo) tag at the N-terminuns to both alleles of XLF, Ku70, or DNA-PKcs and a commercially available proteolysis targeting chimera Halo ligand (Halo-PROTAC) to degrade Halo-tagged proteins in living cells [20,21]. For this study, we also generated U2OS cells expressing Halo-Ku80 from its endogenous locus (Fig. S1a,b). The Halo-DNA-PKcs and Halo-Ku80 cell lines have moderate defects in cell fitness (as evidenced by small colony size, S1c), and the Halo-DNA-PKcs cell strain has a mild NHEJ deficit (Fig. S1d). The Halo-PROTAC efficiently depleted the NHEJ factors as assessed by staining with a fluorescent Halo ligand (JFX650) and western blotting (Fig. 1a,b).

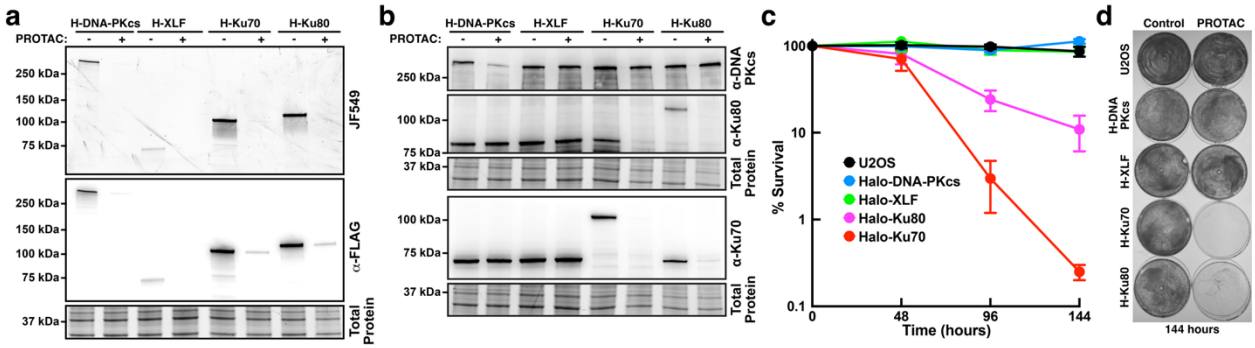


Fig. 1. Cellular ablation of Ku70, or Ku80 but not XLF or DNA-PKcs results in rapid cell death in U2OS cells. a, Fluorescence gel (top) and anti-FLAG western blot (middle), and loading control (bottom) of cell lysates generated from U2OS cells expressing Halo- DNA-PKcs, XLF, Ku70, or Ku80 that had been incubated with 1 μM Halo-PROTAC or vehicle control for 48 hrs**. b,** Western blots probed with antibodies against DNA-PKcs, Ku80, and Ku70 of cell lysates generated from U2OS cells expressing Halo- DNA-PKcs, XLF, Ku70, or Ku80 that had been incubated with 1 μM Halo-PROTAC or vehicle control for 72 hours. **c,** Quantification of MTT cell viability assay of U2OS cells expressing Halo- DNA-PKcs, XLF, Ku70, or Ku80 that had been incubated with 1 μM Halo-PROTAC for the indicated time (N=3 biological replicates, Mean±SEM). **d,** Cell viability analysis using crystal violet staining of U2OS cells expressing Halo- DNA-PKcs, XLF, Ku70, or Ku80 that had been incubated with 1 μM Halo-PROTAC for 6 days.

Depletion of either Halo-Ku70 or Halo-Ku80 results in rapid loss of the other untagged Ku70 or Ku80 subunit, confirming that Ku70 and Ku80 are required to stabilize each other (Fig. 1b).

Incubation of the five U2OS cell lines with 1 μM Halo-PROTAC resulted in significant loss of viability over the course of 6 days for cells expressing Halo-Ku70 or Halo-Ku80 but did not affect viability of parental cells or cells expressing Halo-XLF or Halo-DNA-PKcs (Fig. 1c,d). Halo-Ku70 cells died more rapidly than Halo-Ku80 cells (Fig. 1c, ~3% of Halo-Ku70 or ~24% of Halo-Ku80 are viable after 4 days, and ~0.3% or ~11% after 6 days). Although ~70% of Halo-Ku70 cells are viable after 48 hours exposure to 1 uM Halo-PROTAC, cell viability is severely impacted by exposure to Halo-PROTAC for only 6 hours (and replating in fresh media without PROTAC ligand, Fig. S1e) suggesting that even a temporary depletion of Ku results in severe changes in cellular processes from which cells cannot recover.

Sequence comparisons suggest higher primate specific evolution of the three components of the DNA-PK complex, but not components of the ligase IV complex. To gain insight into possible functions of Ku in higher primates versus other mammals, we compared human NHEJ

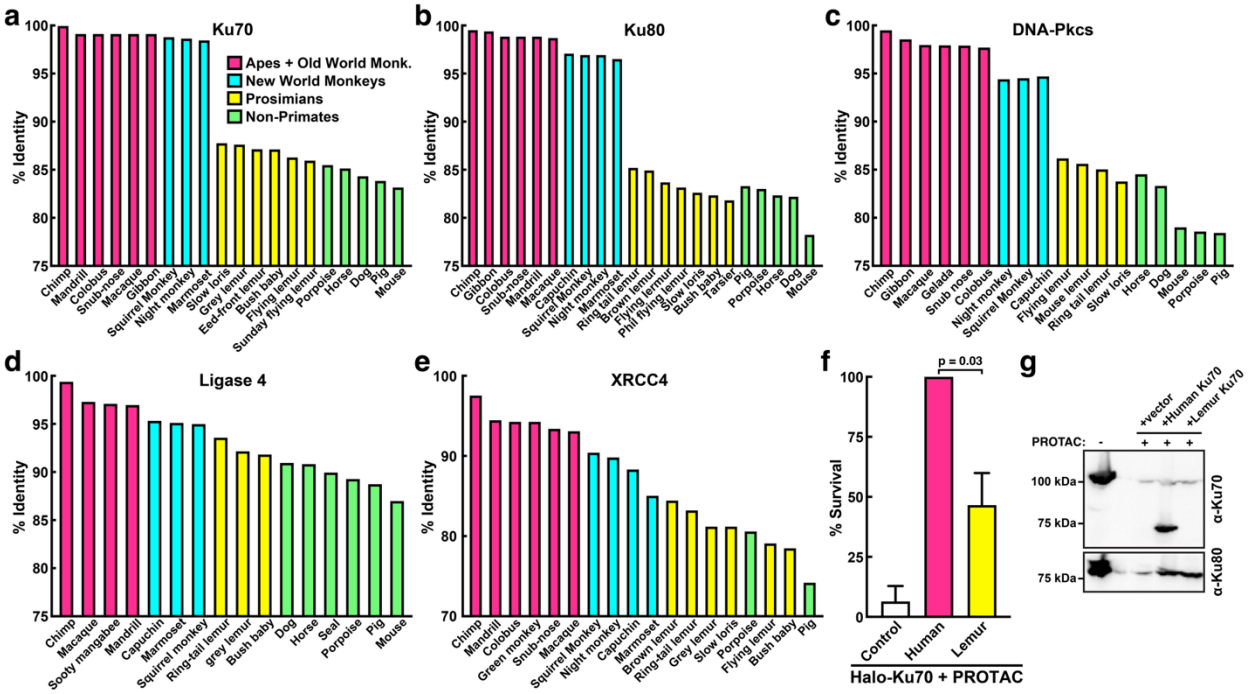


Fig. 2. Distinct evolution of DNA-PK polypeptides in higher primates is required for cell essential function of Ku70 and Ku80. a-e, NHEJ factor orthologues of Ku70 (**a**), Ku80 (**b**), DNA-PKcs (**c**), Lig4 (**d**), or XRCC4 (**e**) from a variety of non-human primates (apes and old-world monkeys, red; new-world monkeys, blue; prosimians, yellow) and other mammals (dog, horse, seal porpoise, pig, and mouse; green) as indicated were compared to the human sequence. Percent identity for each to the human sequence is presented. **f,** Rescue experiments for Halo-Ku70 expressing cells treated with 1 μM Halo-PROTAC, complemented by transient transfection of lemur or human Ku70. Cell survival was assessed by MTT staining and compared to blasticidin resistance conferred by the expression plasmids (N=3 biological replicates, Mean±SEM). **g**, Western blots probing for Ku70 and Ku80 of Halo-Ku70 cells treated with 1 μM Halo-PROTAC, demonstrating that Lemur Ku70 stabilizes human Ku80 to confirm expression of Lemur Ku70 and complex formation with human Ku80.

factor sequences to those of apes and old-world monkeys (red), new-world monkeys (blue), prosimians (yellow), and non-primates (porpoise, horse, dog, pig, mouse, green) (Fig. 2a-e). While sequence changes in XRCC4 and Ligase 4 appear to evolve gradually, sequence identity in the subunits of DNA-PK abruptly changed (>10%) at the prosimian/new world monkey juncture (Fig. 2a-e). These substantial changes in DNA-PK polypeptide sequences, suggest novel properties of Ku70/80 and DNA-PKcs that may be specifically required in higher primates.

Prosimian Ku70 and Ku80 can only partially rescue Halo-Ku depletion in U2OS cells. We next tested whether the higher primate homolog of Ku70 is required to support the cell-essential role of Ku70. To this end, a cDNA encoding Ku70 was isolated from a previously studied [4] cell line from a ring-tailed lemur (prosimian). Expression constructs encoding either human or lemur Ku70 were tested for their ability to rescue cells from endogenous human Halo-Ku70 degradation (Fig. 2f). Lemur Ku70 stabilized human Ku80, confirming that it is expressed and can associate with human Ku80 (Fig. 2g). However, lemur Ku70 expression only partially rescued cell viability after depletion of human Halo-Ku70 (Fig. 2f). We conclude that human Ku70 has specific functional features in higher primates that are important for Ku's essential role in human cells.

Gene expression is significantly altered after 48 hours of Ku70 degradation in U2OS cells, but no activation of an innate immune response is observed. Zhu et al. reported that Ku80 depletion in HCT116 cells dramatically upregulates innate immune signaling via the STAT1 and protein kinase R (PKR) pathways [5]. They propose that Ku functions to restrain abundant Alu RNA molecules by inhibiting their ability to bind to innate immune sensors that recognize viral double-stranded RNA. To investigate the impact of Ku degradation on gene expression in U2OS cells, we carried out RNA-seq experiments comparing control Halo-Ku70 or Halo-DNA-PKcs cells treated with 1 μM Halo-PROTAC for 48 hours, the timepoint immediately prior to rapid cell death of Ku70 depleted cells. No significant changes in gene expression were observed after Halo-DNA-PKcs ablation (Fig. S2a). In contrast, transcription was significantly altered in Halo-Ku70 cells treated with 1 μM Halo-PROTAC compared to untreated controls (Fig. 3a).

We hypothesized that loss or gain of specific transcripts after Ku ablation in diverse cell types might reveal important mechanistic insight into the basis of Ku's essential nature in human cells. After Ku70 degradation in U2OS cells 527 genes were significantly upregulated and 907 genes were significantly downregulated (Fig. 3a-c). When comparing our result to previously published RNAseq data by Zhu *et al*. after Ku80 degradation in HCT116, no strong correlation between genes with altered RNA levels was detected between the two cell types (Fig. 3d). Importantly, we did not observe the upregulation of STAT1 dependent innate immunity genes in U2OS cells,

which was observed in HCT116 cells (e.g. EIF2AK2, Fig. 3a,d). In contrast, key factors involved in double-stranded RNA sensing (e.g. OASL, OAS1), which were found to be upregulated in HCT116 after Ku80 depletion, were significantly downregulated in U2OS after Halo-Ku70 degradation (Fig. 3d). To further analyze the activation of double-stranded RNA sensing in U2OS cells after Halo-Ku70 degradation, we assessed phosphorylation of PKR (EIF2AK2) at T446. Transfection of U2OS cells with double-stranded RNA (polyI:C) but not Halo-Ku70 ablation led to robust induction of PKR phosphorylation (Fig. 3e, Fig. S2b), demonstrating that this double-stranded RNA sensor is not activated in U2OS cells lacking Ku70.

The essential translation initiation factor EIF3B was significantly downregulated in HCT116 and U2OS cells after Ku subunit depletion (Fig. 3a,c ,Fig. S2c). EIF3B protein levels were also reduced in Ku70 depleted HEK293 cells as assessed by mass-spectrometric methods by Kelly et al.[18]. DepMap co-dependency analysis of RNAi experiments revealed that EIF3B, Ku70, and Ku80 are

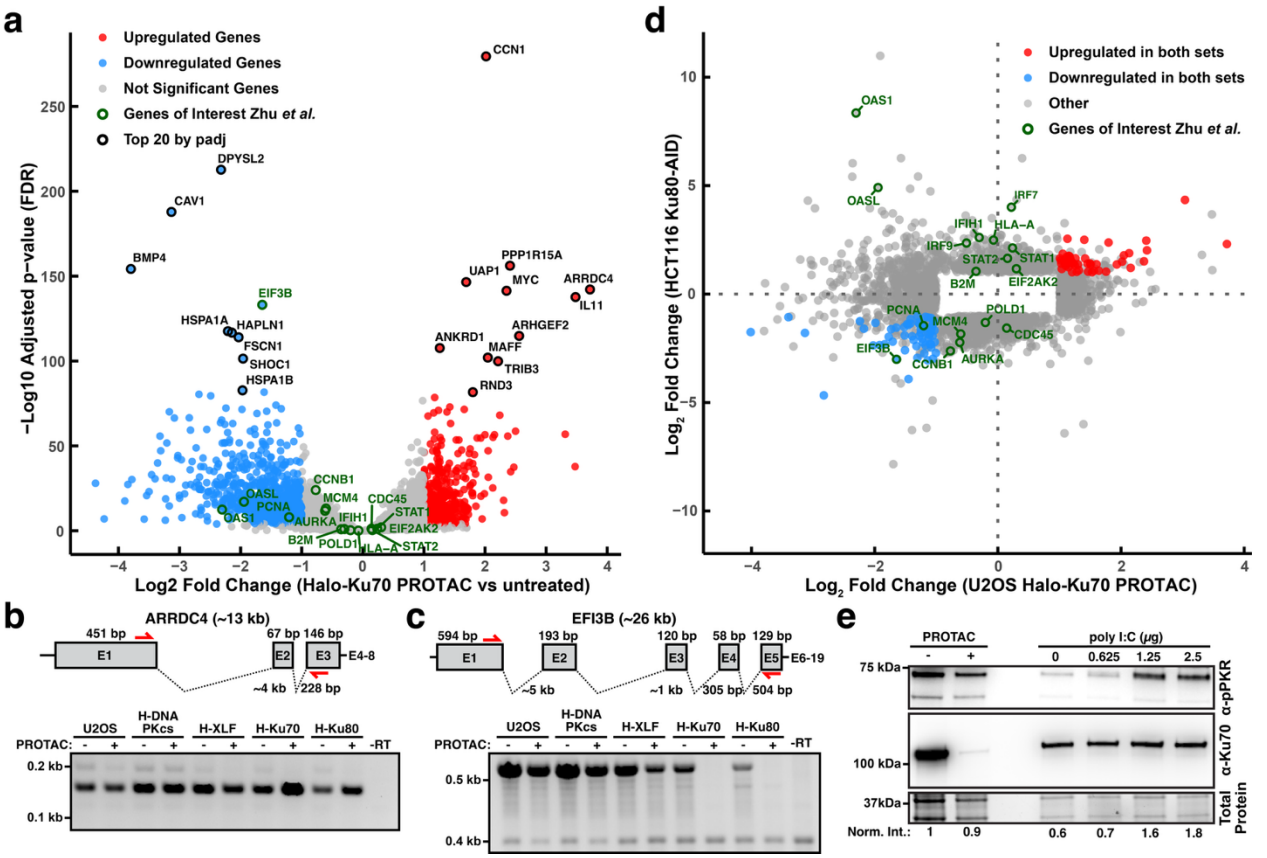


Fig. 3. Gene expression is markedly altered within 48 hours of Ku70 ablation, but no obvious activation of innate immune pathways is observed. a, Volcano plot comparing mRNA levels in untreated Halo-Ku70 cells to cells treated with 1 uM Halo-PROTAC for 48 hours. **b-c,** Validation of upregulation of (**b**) ARRDC4 or downregulation of (**c**) EIF3B in Halo-Ku70 cells treated with Halo-PROTAC by RT-PCR. **d,** Scatter plot of the Log2 fold changes of mRNA levels in U2OS Halo-Ku70 cells and HCT116 Ku80-AID cells after Ku depletion relative to control cells. **e,** Immunoblot demonstrating induction of T446 phosphorylation on PKR by transfection of poly I/C but not by 1 uM Halo-PROTAC treatment of Halo-Ku70 cells. Intensity of pPKR signal was normalized to total protein loading.

highly correlated (Fig. S2c,d), which would be expected if the loss of EIF3B expression is responsible for the reduction in cell viability after Ku degradation. We conclude that activation of innate immune signaling pathways by cytoplasmic double-stranded RNA is unlikely to contribute to cell death in U2OS after degradation of Ku70.

Ablation of either Ku subunit results in dysregulation of RNA splicing. Since we did not observe activation of innate immune signaling in U2OS cell after Ku70 depletion, we considered alternative explanations for the rapid cell death observed. Our RT-PCR validation experiments yielded unexpected products in Ku-ablated cells, potentially resulting from alterations in splicing. Alu elements have been widely implicated in mRNA splicing [22-27], and because Ku was recently demonstrated to specifically bind antisense Alu RNAs [5,28], we analyzed alternative splicing after Ku70 ablation in U2OS cells.

Degradation of Halo-Ku70 in U2OS cells results in significant changes in mRNA splicing (Fig. 4a). These alternative splicing events occur most frequently in mRNAs encoding RNA splicing and RNA processing factors (Fig. 4b). To confirm these findings, we analyzed alternative splicing in RNA-seq data sets after Ku depletion in HCT116 and HEK293 cells published by Zhu. *et al.* [5]. This analysis demonstrated that splicing was also significantly altered in both HCT116 and 293 cells after Ku80 or Ku70 degradation (Fig. S3a-c,h), and the number of alternatively excised introns increased over time (Fig. S3a-c). As in U2OS cells, alternatively spliced genes after Ku depletion in HCT116 and HEK293 were involved in RNA splicing and RNA processing pathways (Fig. S3d-e,i). Overall, 181 genes (enriched for RNA splicing and RNA processing factors) with at least one alternatively spliced intron were shared across the three cell lines ($p<1e-16$), consistent with a common mechanism underlying these changes (Fig. 4c, S3k). Since genes contributing to RNA processing and splicing were enriched in alternatively spliced RNAs in all three cell lines, it is possible that depletion of splicing factors exacerbates the dysregulated splicing that occurs in the absence of Ku. Thus, some observed changes may not be a direct consequence of Ku depletion, per se, but are an indirect consequence of altered protein levels of splicing factors.

To begin to address whether and how splicing dysregulation after Ku ablation is associated with intronic Alu elements, we analyzed introns that were differentially spliced for the presence of Alu elements. In all three cells lines, alternatively spliced introns were highly enriched for SINE sequences, which include Alu elements (Fig. 4d, Fig. S3g,j). To determine the genomic proximity of alternatively spliced introns to Alu elements, we calculated the distance between the midpoint

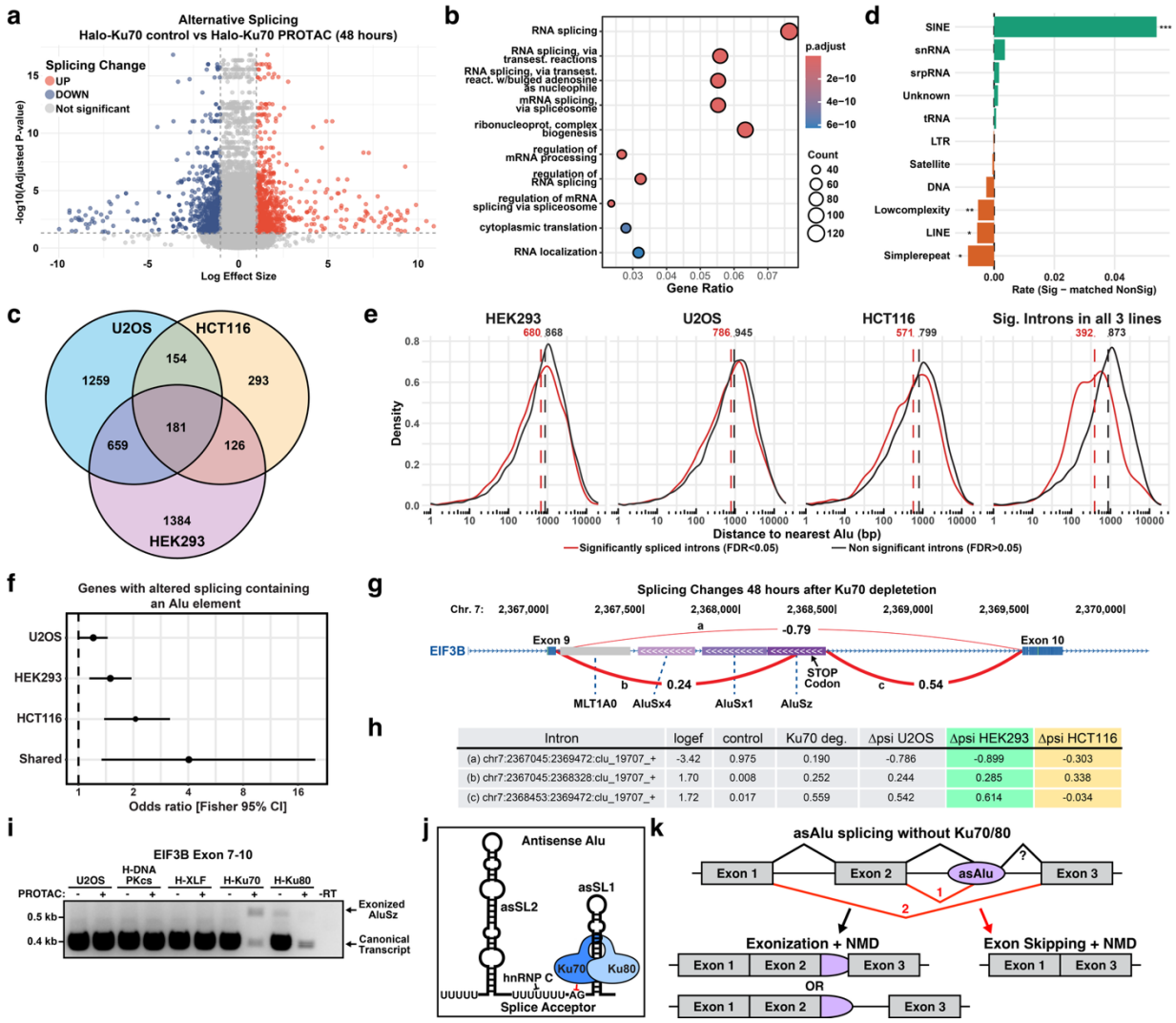


Fig. 4. Ablation of either Ku subunit results in dramatic dysregulation of RNA splicing. a, Volcano plot showing differential splicing events identified by LeafCutter in Ku70-depleted U2OS cells. Each point represents an intron cluster, with significantly up- and downregulated events (|logef| > 1, adjusted P < 0.05) shown in red and blue, respectively; non-significant events are in grey. Dashed lines indicate effect size and significance thresholds. **b,** Gene ontology analysis of biological processed for genes with alternatively spliced introns in U2OS cells 48 hours after Halo-Ku70 depletion. **c,** Venn diagram showing the overlap of genes with altered splicing after Ku depletion in U2OS, HCT116, and HEK293 cells. **d,** Analysis of the enrichment of repetitive elements enriched in alternatively spliced introns in U2OS cells after Halo-Ku70 depletion for 48 hours. **e**, Quantification of the distance between the midpoint of an alternatively spliced (red) or unchanged (black) introns after Ku70 depletion to the closest Alu element in U2OS, HEK293, or HCT116 cells, and for introns with altered splicing in all three cell lines. **f**, Odds ratio of a gene alternatively spliced after Ku70 depletion containing an Alu element. **g**, Diagram of splice junctions in the EIFB3 gene altered after Ku70 depletion in U2OS cells. **h**, Quantification of splice junctions in the EIFB3 gene altered after Ku70 depletion in U2OS, HEK293, and HCT116 cells. **i**, RT-PCR analysis of EIF3B mRNA splicing from exon 9 to 10 after depletion of various NHEJ factors in U2OS cells. **j**, Model of the structure of an antisense Alu element bound by Ku70/80 to block the cryptic splice acceptor site. **k**, Model depicting potential outcomes of Alu element exonization, or exon skipping due to the use of cryptic splice acceptors within antisense Alu elements de-repressed after Ku depletion.

of the affected intron and the nearest annotated Alu element (Fig. 4e). The median distance of alternatively spliced introns to an Alu element was shorter than for introns with unchanged splicing

in the absence of Ku70; this was observed in all three cell lines (Fig. 4e). For the subset of genes that were aberrantly spliced when Ku was ablated in all three cells lines, the distance between the alternatively spliced intron and the closest Alu element was significantly reduced (Fig. 4e). These results indicate that proximity to an Alu element contributes to alternative splicing in the absence of Ku. In support of this interpretation, the odds for an alternatively spliced gene to contain at least one intronic Alu element was higher for genes displaying altered splicing in all three cell lines (Fig. 4f).

It has been previously shown that anti-sense Alu elements impact cryptic splicing more than sense oriented Alu elements [29,30]. Consistent with this observation, the median genomic distance between alternatively spliced introns and antisense Alu repeats was shorter than for sense Alu elements, and this was most evident for introns of alternatively spliced genes shared by the three cell lines (Fig. S4a). To further confirm these findings, we carried out RT-PCR analysis across multiple exon junctions for the *UPF3*, and *MELK* genes (Fig. S4b-e). This RT-PCR analysis also demonstrated altered splicing, including skipping, of exons flanking introns containing antisense Alu elements in Halo-Ku70 depleted U2OS cells.

To define the molecular nature of the alternative splicing events, we focused on selected novel splice junctions that are upregulated after Ku70 depletion. The essential EIF3B gene, which is downregulated in all three cell lines, contains two novel splice junctions between exons 9 and 10, which are upregulated after Ku70 depletion at the expense of the canonical junction that properly connects these exons (Fig. 4g,h). The splice junctions upregulated after Ku70 depletion in U2OS cells contain a splice acceptor and donor within an antisense Alu element in the intron intervening exons 9 and 10 (Fig. 4g,h). Precisely the same altered splice junctions within this Alu element were also detected in HCT116 and HEK293 cells after Ku depletion (Fig. 4h). This alternative splicing event was confirmed by RT-PCR (Fig. 4i). A small amount of the transcript containing the exonized Alu element was also detected in control Halo-Ku80 cells (Fig. 4i), suggesting that the HaloTag on the N-terminus of Ku80 slightly impacts its ability to limit aberrant splicing. All other splice junctions within the EIF3B transcript were unchanged after Ku70 depletion (Fig. S4f) notwithstanding the presence of intronic antisense Alu repeats; this demonstrates that the altered splicing caused by Ku depletion is specific to certain introns, rather than a general defect in mRNA processing. To expand this analysis, we assessed whether splice junctions upregulated after Ku70 depletion had start or end points that mapped within Alu elements. In genes that were alternatively spliced in all three cell lines, splice junctions upregulated upon Ku depletion showed a strong likelihood of having start or endpoints overlapping with the genomic coordinates of

intronic Alu elements (Fig. S4g). These Alu repeats were almost exclusively in antisense orientation to the host gene (Fig. S4h), consistent with the observation that antisense Alu elements impact cryptic splicing more than sense oriented

Antisense Alu elements typically contain two stem loops preceded by poly-uridine stretches (Fig. 4j), which can be bound by U2AF2 in the context of splice acceptor recognition [31]. To better understand how Ku prevents the use of the cryptic splice acceptor in intron 9 of the EIF3B gene, we predicted the secondary structure of the antisense AluSz element containing the splice acceptor (Fig. S5). This analysis revealed that the cryptic splice acceptor is located immediately upstream of the asSL1 of the antisense Alu element, which was shown to be the primary binding site of Ku [5], suggesting that Ku70/80 binding to this stem loop directly interferes with splicing, potentially by sterically preventing the association of the splicing machinery. Importantly, we identified additional examples of cryptic splice acceptors upregulated after Ku depletion in all 3 cell lines immediately upstream of the asSL1 stem loop in antisense Alu elements in the *RPL27A* and *EXOSC9* genes (Fig. S6a,b). Furthermore, we also observed cryptic splice acceptors in all three cell lines with increased use after Ku depletion upstream of asSL2 of antisense Alu elements (Fig. S6c-d), which is also preceded by a poly-uridine stretch (Fig. 4j) and was shown to associate with Ku [5]. Together, these results demonstrate that intronic antisense Alu elements contain cryptic splice acceptors immediately upstream of Ku binding stem loops that are activated after Ku70 depletion.

Discussion

Alu invasion of primate genomes is a double-edged sword. It is well-appreciated that the NAHR promoted by Alu elements functioned to diversify primate genomes allowing for the evolution of traits like full color vision, advanced neuronal development and function (reviewed in [32], lack of tails in apes [33] etc.). It is similarly well-documented that detrimental Alu-mediated recombination events are the basis of numerous genetic diseases (for example hemophilia, familial hypercholesterolemia, spinal muscular atrophy, etc.) [34,35]. Moreover, Alu-mediated somatic recombination events promote many types of human cancers (in genes like BRCA1, BRCA2, MSH2, MLL1) (reviewed in [14,36]). On the other hand, the presence of Alu elements in transcribed mRNAs further diversifies the potential transcriptome by providing a robust mechanism to alter protein products by mRNA splice variation, as has been appreciated for several decades [22-27].

But it appears that this acquired ability to diversify protein expression by insertion of Alu elements in protein coding genes is apparently too robust – necessitating co-evolution of a mechanism to regulate splice variation. The expansion of Alu elements in primate genomes resulted in large increases in the cellular levels of the SRP Alu-binding heterodimer SRP9/14 as well as components of the DNA-PK complex [4]. Although other explanations are possible, the simplest model to explain how Ku (with or without DNA-PKcs) regulates splicing, is that Ku (with its relatively weak dsRNA binding activity) binds intronic Alu elements in immature RNA transcripts, modulating potential alternative splice sites. Consistent with this model, Ku has been demonstrated previously to associate with both RNA and RNA processing factors [37-39], with the spliceosome [40,41], and cellular Ku/RNA interactions are predominately (98%) interactions of Ku with antisense Alu elements in mRNA transcripts [5].

Our data presented here demonstrate that Ku depletion leads to activation of cryptic splice acceptors immediately upstream of the RNA stem loops within antisense Alu elements that Ku has been shown to bind [5]. We observed both exonization of Alu elements and exon skipping in close proximity to Alu sequences (Fig. 4k). Splice acceptors contain two critical sequence elements, a poly-pyrimidine tract followed by the essential AG dinucleotide, which marks the end of the intron [42]. Both of these elements are present in antisense but not sense Alu elements. The poly-pyrimidine tract is bound by the splicing factor U2AF2 [42], and elegant work has demonstrated that hnRNPC competes with U2AF2 to prevent the usage of cryptic splice acceptors within antisense Alu elements [31]. The AG dinucleotide is recognized by U2AF1, which also associates with the first two nucleotides of the subsequent exon [43]. Strikingly, the AG dinucleotide of splice acceptors upregulated after Ku degradation are located within a few nucleotides of the Ku70/80

binding stem loops. We therefore propose a model in which Ku70/80 binding directly competes with U2AF1 binding to the splice acceptor within antisense Alu elements (Fig. 4j). Exonization of antisense Alu elements during evolution therefore likely requires mutations that disrupt stem loop formation and sequence changes that weaken hnRNPC binding to the poly-pyrimidine of the splice acceptor site, providing multiple layers of regulation to buffer potentially detrimental impacts of the expansion Alu elements in the genomes of higher primates.

Importantly, the dysregulation of splicing after Ku depletion was documented in three independent cell systems using distinct methodological approaches targeting both Ku subunits. The genes impacted by alternative splicing include factors that participate in essential processes such as translation, ribosome biogenesis, and splicing itself, which likely explains the rapid cell death observed after Ku depletion. It is possible that some of the observed changes in RNA splicing are indirect effects, resulting from expression changes in splicing factors. However, it is important to note that all data included in our analysis stemmed from the earliest available time points, before substantial cell death was observed in U2OS and HCT116 cells, which should minimize secondary effects. The core set of genes aberrantly spliced after Ku depletion in all three cell lines due to antisense Alu elements includes a number of essential genes (e.g. *EIF3B*, *EXOSC9*, *RPL27a*, *PDCD11*). The loss of any of these proteins could in principle result in cell death and due to the broad effect of Ku depletion on splicing it is likely not a single factor that is solely responsible for the observed phenotype. It is possible that innate immune signaling triggered by Alu elements contributes to cell death in HCT116 cells. However, we did not observe activation of this pathway in U2OS cells and in HCT116 cells disrupting innate immune signaling only partially rescues cell viability [5]. The preference of Ku for antisense Alu elements is consistent with its role in splicing [5], considering that only these sequences contain the poly-pyrimidine tracts that are key elements of splice acceptors. In contrast, it is not immediately clear how preferentially binding to antisense Alu elements would prevent dsRNA mediated innate immune signaling triggered by sense Alu sequences.

In total, the results presented here strongly support a model in which DNA-PK (Ku70, Ku80, DNA-PKcs) expression increased in unison with the expansion of Alu elements in the genomes of higher primates to inhibit splice acceptor sites within antisense Alu sequences. We propose that Ku acts as a critical regulator of splicing homeostasis in primate genomes, balancing the pervasive influence of Alu elements on transcript diversity while safeguarding the fidelity of RNA maturation.

Methods

Construction of HF-tagged U2OS cell strains. Halo-Ku70, Halo-XLF, and Halo-DNA-PKcs cell strains have been described previously[20]. Generation of Halo-Ku80 cells was accomplished via the same methods. Cells with Halo tags incorporated at the N-terminus were screened first via genomic DNA-PCR and then confirmed by western blot analyses as described previously [20].

Cell viability assay. MTT staining was performed to assess cell viability for 4-day survival assay as well as 96 hours survival time-course of halo-tagged cell lines. Duplicate wells of each cell line were plated into 24-well plates, containing medium with DMSO or 1 μM HaloPROTAC-E (AOBIOUS) (CAS: 2365478-58-4). After 4 days of Halo-Protac treatment (or at 24 h, 48 h, 72 h, 96 h time points), cells were treated with 1 mg/ml MTT (GoldBio) solution for 1 h. Medium containing MTT was then removed and 200 μL DMSO was added per well to lyse cells and solubilize the formazan crystals formed. Absorbance was read at 570 nm to determine relative survival. Background reading was collected from blank wells without cells following the same treatment as experiment and control wells.

Colony forming assay. Clonogenic survival assays were performed in U2OS cells. Briefly, two hundred cells were plated for each transfectant into complete medium containing the indicated dose of calicheamycin in 60 mm diameter tissue culture dishes. After 8 days, cell colonies were stained with 1% (w/v) crystal violet in ethanol to measure relative survival. Four independent experiments were performed, and results were averaged.

RNA isolation and RT-PCR. Total cellular RNA was isolated using Trizol (Invitrogen) according to the manufacture's recommendations. cDNA was synthesized using Superscript-RT (ThermoFischer) according to the manufacture's recommendations. Resultant cDNA was used in RT-PCR experiments using Go-Taq (ThermoFischer). RT-PCR primers are listed in supplemental Table 2.

Western blotting. SDS-PAGE and western transfer were carried out using standard protocols and probed using primary antibodies targeting Ku70 (1:1000, Neomarkers, Ab4), Ku80 (1:1000, XYC), DNA-PKcs (1:1000, 42-27, a generous gift from Tim Carter), phospho-T446 PKR (1:1000, Invitrogen, 111), HRP-FLAG-M2 (1:4000, Sigma Aldrich A8592), and HRP conjugated anti-mouse and anti-rabbit secondary antibodies. For fluorescent HaloTag staining, cells were incubated for 30 minutes with 100 nM of JF549-HaloTag ligand prior to generating cell lysates.

RNA-seq. The Illumina Stranded mRNA Library Prep kit (Illumina, Cat no. 20040534) with IDT for Illumina RNA Unique Dual Index adapters was used for library preparation following the manufacturer’s recommendations but using half-volume reactions. Qubit™ dsDNA HS

(ThermoFischer Scientific, Cat no. Q32851) and Agilent 4200 TapeStation HS DNA1000 assays (Agilent, Cat no. 5067-5584) were used to measure quality and quantity of the generated libraries. The libraries were pooled in equimolar amounts, and the Invitrogen Collibri Quantification qPCR kit (Invitrogen, Cat no. A38524100) was used to quantify the pooled library. The pool was loaded onto 2 lanes of a NovaSeq S4 flow cell at Genewiz Sequencing, and sequencing was performed in a 2x150 bp paired end format using a NovaSeq 6000 v1.5 100-cycle reagent kit (Illumina, Cat no. 20028316). Base calling was performed with Illumina Real Time Analysis (RTA; v3.4.4), and the output of RTA was demultiplexed and converted to the FastQ format with Illumina Bcl2fastq (v2.20.0).

RNA-seq analysis. RNA sequencing reads were analyzed using the Michigan State University High Performance Computing Center using a custom pipeline built in Snakemake (Version 7.32.4) (https://github.com/kaylaconner/olivelab-rnaseq/tree/main)[44]. Read quality was assessed using FastQC (Version 0.12.1) [45]. Read mapping was performed with Bowtie2 (Version 2.5.1) using default settings and the Hg19 genome. Aligned reads were counted using the featureCounts program in Subread (Version 2.0.6) [46]. Normalized count tables and differential gene expression analysis was done using the DESeq2 package (Version 1.40.2) in R (Version 4.3.0) [47]. Counts were pre-filtered to only keep genes that had >10 counts in at least three samples. Publicly available RNA-seq data from HCT116 and HEK293 cells were obtained from the Gene Expression Omnibus (GEO) accession GSE294709.

Alignment and annotation of differentially spliced introns. For all datasets, Fastq files were trimmed with trimgalore (https://github.com/FelixKrueger/TrimGalore) and aligned on GRCh38 with STAR (v2.7.9a), using parameters: --twopassMode Basic --outSAMstrandField intronMotif --outSAMtype BAM SortedByCoordinate --outSAMunmapped Within --outSAMattributes Standard --sjdbGTFfile /home/giovanni/UCSC_TABLES/hg38/gencode.v46.annotation.gtf --sjdbOverhang 100. Differential splicing was quantified using LeafCutter (v0.2.9) with default parameters, which clusters spliced reads based on shared intron–exon boundaries and tests for differential intron excision ratios between experimental and control samples. Aligned RNA-seq reads in BAM format were first processed to extract splice junctions using the leafcutter_cluster_regtools.py utility. Only junctions supported by at least 30 split reads across all samples and detected in ≥3 biological replicates were retained for analysis. Each intron cluster was tested for differential splicing using the Dirichlet-multinomial model implemented in LeafCutter's leafcutter_ds.R script. Log-likelihood ratio tests (LRT) were used to identify clusters with significant changes in intron usage between knockout and control conditions. P-values were adjusted for multiple testing using the Benjamini–

Hochberg false discovery rate (FDR) correction, and clusters with FDR ≤ 0.05 were considered significantly differentially spliced. For each significant cluster, effect sizes were computed as the difference in intron inclusion levels (ΔPSI) between conditions. Clusters were further classified as Up or Down based on the sign of ΔPSI, indicating increased or decreased intron usage in the knockout condition, respectively. Gene-level associations were determined using GENCODE v48 annotations, assigning each cluster to the host gene(s) overlapping its intronic coordinates. To integrate splicing data with repeat element annotations, introns were intersected with RepeatMasker annotations (GRCh38, UCSC RepeatMasker track) using BEDTools (v2.27.0). This allowed annotation of introns for the presence and orientation of repeat elements relative to the transcriptional strand.

Gene Ontology analysis. Lists of significantly spliced genes (FDR < 0.05) identified by LeafCutter in each dataset were analyzed for functional enrichment using the Biological Process (BP) category of the Gene Ontology. Enrichment analysis was performed in R with the clusterProfiler package (v4.8.1), using org.Hs.eg.db as the annotation database. The universe gene set consisted of all genes annotated in org.Hs.eg.db. P-values were adjusted for multiple testing using the Benjamini–Hochberg method, and the top significantly enriched terms (p.adjust < 0.05) were visualized as dot plots ranked by adjusted P-value and gene ratio.

Matched random permutation analysis. Tables were constructed containing, for each intron, its genomic coordinates, host gene ID, repeat content, GC percentage, read depth, ΔPSI, and differential splicing significance. To assess whether significantly spliced introns were enriched in repetitive elements relative to expectation, introns belonging to significantly spliced clusters (FDR < 0.05) were compared to an equal number of non-significant introns randomly sampled from the same dataset. Random selection was repeated 10,000 times, with matching on key genomic covariates including intron length, GC content, expression level of the host gene, and mappability, to control for potential confounders. At each iteration, we calculated the proportion of introns overlapping the repeat class of interest (e.g., Alu, LINE, SINE) in both sets, and computed the difference in overlap rate (Δ = − rate_{matched}) [48]. The empirical distribution of Δ under random matching was used to estimate a two-tailed permutation P-value, representing the probability of observing an enrichment (or depletion) as extreme as the observed Δ under the null hypothesis of no association. False discovery rates (FDR) were obtained by Benjamini–Hochberg correction across repeat classes. Median enrichment and 95% empirical confidence intervals were derived from the resampled distributions.

Odds ratio enrichment analysis. To quantify the enrichment of Alu elements among genes showing significant alternative splicing, we performed a Fisher's exact test–based odds ratio analysis. For each dataset, we first identified genes containing at least one significantly spliced intron (FDR < 0.05) using LeafCutter. Genes were classified as Alu-positive if at least one Alu element, as annotated by RepeatMasker (UCSC hg38), overlapped an intron within their GENCODE v48 gene body. The analysis universe consisted of all genes tested by LeafCutter in each dataset. For each dataset, a 2×2 contingency table was constructed comparing the number of genes with or without intronic Alu elements between the significant and non-significant gene groups. Odds ratio (OR) and 95% confidence interval (CI) were computed using Fisher's exact test.

Distance analysis between introns and Alu elements. To evaluate the spatial relationship between alternatively spliced introns and nearby Alu elements, we calculated the genomic distance from each intron midpoint to its nearest Alu repeat using the UCSC hg38 RepeatMasker annotation. Intron coordinates were derived from LeafCutter effect size tables and merged with corresponding cluster significance results to classify introns as significant (FDR < 0.05) or non-significant. The midpoint of each intron ([(start + end)/2]) was used to represent its genomic position, and distances to the nearest Alu element within a 1bp-20Kb distance range were computed with the distanceToNearest() function from the GenomicRanges R package (v1.54). Distances were calculated genome-wide, ignoring strand, and stored both as raw base-pair values and in $\log_{10}$ scale for visualization. The distributions of intron–Alu distances were compared between significant and non-significant introns using the Wilcoxon rank-sum test, and effect sizes were estimated using Cliff's delta. For the multi-dataset analysis, introns that were significantly spliced in all three cell lines were intersected by exact genomic coordinates and analyzed as a combined subset against all non-significant introns in the Leafcutter datasets of the three cell lines.

Analysis of Alu enrichment and Alu orientation bias at non-canonical splice junctions. LeafCutter output tables were used to identify significantly altered introns (FDR ≤ 0.05) in each dataset. For each intron cluster, splice junctions with positive ΔPSI (ΔPSI > 0) were defined as newly gained (non-canonical) splice junctions upon Ku depletion. To ensure that only truly novel junctions were considered, we removed any ΔPSI>0 splice-site endpoints that overlapped with canonical junction endpoints (ΔPSI<0) from the same cluster. The genomic coordinates of the remaining ΔPSI>0 junction endpoints were intersected with RepeatMasker annotations to identify those overlapping Alu elements. For each dataset and for the set of reproducible junctions shared

across all three datasets (≥10 bp intron overlap), we quantified Alu enrichment relative to a background expectation derived from the intronic Alu content of all significantly spliced introns. Expected endpoint overlap with Alu was calculated as the sum of Alu base-pair coverage across these introns, weighted by their lengths. Statistical significance was assessed using a binomial test (observed vs expected overlap) and a permutation test (10,000 random endpoint placements within the same introns). To assess orientation bias of Alu-overlapping non-canonical junctions, non-canonical junction endpoints (ΔPSI>0 and filtered as above) were intersected with Alu elements to determine their overlap. Each Alu hit was assigned a strand orientation relative to the host gene by intersecting endpoints with GENCODE v48 gene annotations. Alu elements were classified as “sense” if the RepeatMasker strand matched the gene strand, and “antisense” otherwise. Orientation bias was evaluated for each dataset and for the reproducible 3-dataset intron set using a two-sided binomial test (H_0: 50% sense / 50% antisense).

All analyses were performed in R (v4.4.2) and plots were generated with the ggplot2 package.

RNA structure prediction. RNA structure models were generated using the RNAfold webserver using the minimal free energy approach.

Acknowledgements

We thank our colleagues Kefei Yu and Yu Zhang for many helpful suggestions and careful review of the manuscript. This work was supported by NIH grant (DP2 GM142307) to J. C. S. and by the USDA National Institute of Food and Agriculture (1019208), and Public Health Service (AI048758, AI147634 to K.M.) and the Meek/Hasemann Cancer Research Fund.

Author contributions

Conceptualization: G.P., J.C.S., and K.M.; Experiments: M.M., J.C.S., Ga.P., J.R.H., T.J., and K.M. Data Analysis: G.P., Ga.P., A.O., J.G., J.C.S., K.M.; Writing: Original Draft: K.M.; Writing: Review and Editing: G.P., J.C.S., and K.M.

Data Availability

RNA-seq data have been submitted to the gene expression omnibus (GEO). All other data is available from J.C.S. and K.M. upon reasonable request.

Competing Interests

The authors declare no competing interests.

References

1 Ule, J. Alu elements: at the crossroads between disease and evolution. *Biochemical Society transactions* **41**, 1532-1535 (2013). https://doi.org:10.1042/BST20130157

2 Deininger, P. Alu elements: know the SINEs. *Genome biology* **12**, 236 (2011). https://doi.org:10.1186/gb-2011-12-12-236

3 Bourque, G. *et al.* Ten things you should know about transposable elements. *Genome biology* **19**, 199 (2018). https://doi.org:10.1186/s13059-018-1577-z

4 Pascarella, G. *et al.* Compared to other NHEJ factors, DNA-PK protein and RNA levels are markedly increased in all higher primates, but not in prosimians or other mammals. *DNA repair* **142**, 103737 (2024). https://doi.org:10.1016/j.dnarep.2024.103737

5 Zhu, Y. *et al.* Ku limits RNA-induced innate immunity to allow Alu expansion in primates. *Nature* **643**, 562-571 (2025). https://doi.org:10.1038/s41586-025-09104-w

6 Meek, K., Gupta, S., Ramsden, D. A. & Lees-Miller, S. P. The DNA-dependent protein kinase: the director at the end. *Immunological reviews* **200**, 132-141 (2004). https://doi.org:10.1111/j.0105-2896.2004.00162.x

7 Buehl, C. J. *et al.* Two distinct long-range synaptic complexes promote different aspects of end processing prior to repair of DNA breaks by non-homologous end joining. *Molecular cell* (2023). https://doi.org:10.1016/j.molcel.2023.01.012

8 Goff, N. J., Mikhova, M., Schmidt, J. C. & Meek, K. DNA-PK: A synopsis beyond synapsis. *DNA repair* **141**, 103716 (2024). https://doi.org:10.1016/j.dnarep.2024.103716

9 Chaplin, A. K. *et al.* Cryo-EM of NHEJ supercomplexes provides insights into DNA repair. *Molecular cell* **81**, 3400-3409 e3403 (2021). https://doi.org:10.1016/j.molcel.2021.07.005

10 Meek, K. *et al.* SCID in Jack Russell terriers: a new animal model of DNA-PKcs deficiency. *Journal of immunology* **167**, 2142-2150 (2001).

11 Finnie, N. J., Gottlieb, T. M., Blunt, T., Jeggo, P. A. & Jackson, S. P. DNA-dependent protein kinase activity is absent in xrs-6 cells: implications for site-specific recombination and DNA double-strand break repair. *Proceedings of the National Academy of Sciences of the United States of America* **92**, 320-324 (1995).

12 Pascarella, G. *et al.* Recombination of repeat elements generates somatic complexity in human genomes. *Cell* **185**, 3025-3040 e3026 (2022). https://doi.org:10.1016/j.cell.2022.06.032

13 Stankiewicz, P. & Lupski, J. R. Genome architecture, rearrangements and genomic disorders. *Trends in genetics : TIG* **18**, 74-82 (2002). https://doi.org:10.1016/s0168-9525(02)02592-1

14 Song, X. *et al.* Predicting human genes susceptible to genomic instability associated with Alu/Alu-mediated rearrangements. *Genome research* **28**, 1228-1242 (2018). https://doi.org:10.1101/gr.229401.117

15 Li, G., Nelsen, C. & Hendrickson, E. A. Ku86 is essential in human somatic cells. *Proceedings of the National Academy of Sciences of the United States of America* **99**, 832-837 (2002). https://doi.org:10.1073/pnas.022649699

16 Fattah, F. J., Lichter, N. F., Fattah, K. R., Oh, S. & Hendrickson, E. A. Ku70, an essential gene, modulates the frequency of rAAV-mediated gene targeting in human somatic cells. *Proceedings of the National Academy of Sciences of the United States of America* **105**, 8703-8708 (2008). https://doi.org:10.1073/pnas.0712060105

17 Woodbine, L. *et al.* PRKDC mutations in a SCID patient with profound neurological abnormalities. *The Journal of clinical investigation* **123**, 2969-2980 (2013). https://doi.org:10.1172/JCI67349

18 Kelly, R. D. *et al.* Noncanonical functions of Ku may underlie essentiality in human cells. *Sci Rep* **13**, 12162 (2023). https://doi.org:10.1038/s41598-023-39166-7

19 Wang, Y., Ghosh, G. & Hendrickson, E. A. Ku86 represses lethal telomere deletion events in human somatic cells. *Proceedings of the National Academy of Sciences of the United States of America* **106**, 12430-12435 (2009). https://doi.org:10.1073/pnas.0903362106
20 Mikhova, M. *et al.* Single-molecule imaging reveals the kinetics of non-homologous end-joining in living cells. *Nature communications* **15**, 10159 (2024). https://doi.org:10.1038/s41467-024-54545-y
21 Caine, E. A. *et al.* Targeted Protein Degradation Phenotypic Studies Using HaloTag CRISPR/Cas9 Endogenous Tagging Coupled with HaloPROTAC3. *Curr Protoc Pharmacol* **91**, e81 (2020). https://doi.org:10.1002/cpph.81
22 Ponicsan, S. L., Kugel, J. F. & Goodrich, J. A. Genomic gems: SINE RNAs regulate mRNA production. *Current opinion in genetics & development* **20**, 149-155 (2010). https://doi.org:10.1016/j.gde.2010.01.004
23 Walters, R. D., Kugel, J. F. & Goodrich, J. A. InvAluable junk: the cellular impact and function of Alu and B2 RNAs. *IUBMB Life* **61**, 831-837 (2009). https://doi.org:10.1002/iub.227
24 Hasler, J. & Strub, K. Alu elements as regulators of gene expression. *Nucleic acids research* **34**, 5491-5497 (2006). https://doi.org:10.1093/nar/gkl706
25 Nakama, M. *et al.* Intronic antisense Alu elements have a negative splicing effect on the inclusion of adjacent downstream exons. *Gene* **664**, 84-89 (2018). https://doi.org:10.1016/j.gene.2018.04.064
26 Nakama, M., Imanaka, B. & Kimoto, Y. Intrinsic Alu affects for RNA splicing in a minigene model. *Biochem Biophys Rep* **42**, 102002 (2025). https://doi.org:10.1016/j.bbrep.2025.102002
27 Daniel, C., Behm, M. & Ohman, M. The role of Alu elements in the cis-regulation of RNA processing. *Cellular and molecular life sciences : CMLS* **72**, 4063-4076 (2015). https://doi.org:10.1007/s00018-015-1990-3
28 Shao, Z. *et al.* DNA-PKcs has KU-dependent function in rRNA processing and haematopoiesis. *Nature* **579**, 291-296 (2020). https://doi.org:10.1038/s41586-020-2041-2
29 Makalowski, W., Mitchell, G. A. & Labuda, D. Alu sequences in the coding regions of mRNA: a source of protein variability. *Trends in genetics : TIG* **10**, 188-193 (1994). https://doi.org:10.1016/0168-9525(94)90254-2
30 Sorek, R., Ast, G. & Graur, D. Alu-containing exons are alternatively spliced. *Genome research* **12**, 1060-1067 (2002). https://doi.org:10.1101/gr.229302
31 Zarnack, K. *et al.* Direct competition between hnRNP C and U2AF65 protects the transcriptome from the exonization of Alu elements. *Cell* **152**, 453-466 (2013). https://doi.org:10.1016/j.cell.2012.12.023
32 Wang, J., Weatheritt, R. & Voineagu, I. Alu-minating the Mechanisms Underlying Primate Cortex Evolution. *Biol Psychiatry* **92**, 760-771 (2022). https://doi.org:10.1016/j.biopsych.2022.04.021
33 Xia, B. *et al.* On the genetic basis of tail-loss evolution in humans and apes. *Nature* **626**, 1042-1048 (2024). https://doi.org:10.1038/s41586-024-07095-8
34 Sen, S. K. *et al.* Human genomic deletions mediated by recombination between Alu elements. *American journal of human genetics* **79**, 41-53 (2006). https://doi.org:10.1086/504600
35 Hedges, D. J. & Deininger, P. L. Inviting instability: Transposable elements, double-strand breaks, and the maintenance of genome integrity. *Mutation research* **616**, 46-59 (2007). https://doi.org:10.1016/j.mrfmmm.2006.11.021
36 Ade, C., Roy-Engel, A. M. & Deininger, P. L. Alu elements: an intrinsic source of human genome instability. *Curr Opin Virol* **3**, 639-645 (2013). https://doi.org:10.1016/j.coviro.2013.09.002

37 Adelmant, G. *et al.* DNA ends alter the molecular composition and localization of Ku multicomponent complexes. *Molecular & cellular proteomics : MCP* **11**, 411-421 (2012). https://doi.org:10.1074/mcp.M111.013581

38 Dvir, A., Peterson, S. R., Knuth, M. W., Lu, H. & Dynan, W. S. Ku autoantigen is the regulatory component of a template-associated protein kinase that phosphorylates RNA polymerase II. *Proceedings of the National Academy of Sciences of the United States of America* **89**, 11920-11924 (1992).

39 Yoo, S. & Dynan, W. S. Characterization of the RNA binding properties of Ku protein. *Biochemistry* **37**, 1336-1343 (1998). https://doi.org:10.1021/bi972100w

40 Jurica, M. S. & Moore, M. J. Pre-mRNA splicing: awash in a sea of proteins. *Molecular cell* **12**, 5-14 (2003). https://doi.org:10.1016/s1097-2765(03)00270-3

41 Jurica, M. S., Licklider, L. J., Gygi, S. R., Grigorieff, N. & Moore, M. J. Purification and characterization of native spliceosomes suitable for three-dimensional structural analysis. *RNA* **8**, 426-439 (2002). https://doi.org:10.1017/s1355838202021088

42 Wilkinson, M. E., Charenton, C. & Nagai, K. RNA Splicing by the Spliceosome. *Annual review of biochemistry* **89**, 359-388 (2020). https://doi.org:10.1146/annurev-biochem-091719-064225

43 Yoshida, H. *et al.* Elucidation of the aberrant 3' splice site selection by cancer-associated mutations on the U2AF1. *Nature communications* **11**, 4744 (2020). https://doi.org:10.1038/s41467-020-18559-6

44 Koster, J. & Rahmann, S. Snakemake-a scalable bioinformatics workflow engine. *Bioinformatics* **34**, 3600 (2018). https://doi.org:10.1093/bioinformatics/bty350

45 andrews, S. FastQC: A quality control tool for high throughput sequence data. . *Available from: http://www.bioinformatics.babraham.ac.uk/projects/fastqc/.* (2010).

46 Liao, Y., Smyth, G. K. & Shi, W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923-930 (2014). https://doi.org:10.1093/bioinformatics/btt656

47 Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**, 550 (2014). https://doi.org:10.1186/s13059-014-0550-8

48 Wensing, T., van Gent, D. M., Schotman, A. J. & Kroneman, J. Hyperlipoproteinaemia in ponies: mechanisms and response to therapy. *Clinica chimica acta; international journal of clinical chemistry* **58**, 1-15 (1975).

Supplemental Data

Supplementary table 1. Counts of Leafcutter differential splicing results, Alu enrichment in all datasets and Alu annotation in GENCODE v48.

Supplemental Table 2.

Primers used for RT-PCR

	FORWARD	REVERSE
EFI3B e1-e5	AGGAAACGAGGGCAGAGCG	CCAGAATATGGAAGTGCGGT
EFI3B e7-e10	GACCCTCAGGCCATAATCAT	GCCTTTCGGAGTCCTATCTA
ARRDC4 e1-e3	TCTCGGAGGTGGAGTACCTG	CCCAGTAAACGAGGTGACCAAA
MELK e1-e4	CCGGCGATCGAAAAGATTCT	TCAGGTTCTTCAAGGCCTCA
MELK e6-e12	GATTGACTTTGGTCTCTGTGC	AGCACTGGCTTGTCCACAGG
UPF3 e6-e9	GAAGAACCACACCTCTTTTGG	TCTCTCCACATCCCTGTGCT

Supplemental Figure 1.

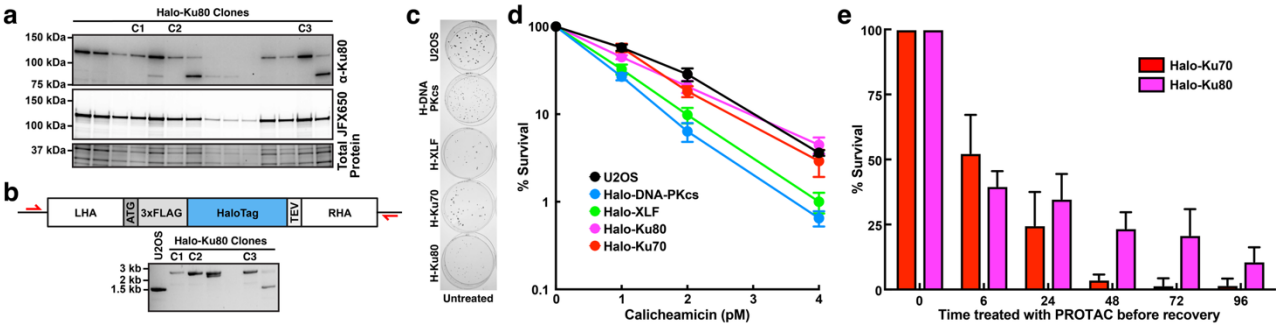


Figure S1. HF-DNA-PKcs and HF-Ku80 have modest fitness defects; HF-XLF and XF-DNA-PKcs have slightly reduced calicheamicin resistance compared to wild U2OS cells. a, Anti-Ku80 western blot (top), fluorescence gel (middle) and loading control (bottom) of cell lysates generated from U2OS cells expressing Halo-Ku80. **b,** Cartoon depicting Halo-tag targeting construct and primer positions used to assess targeting as shown in PCR screen (bottom panel). **c,** U2OS cells were plated at cloning density and stained with crystal violet 8 days later. Colony size is markedly reduced in HF-DNA-PKcs and HF-Ku80 cells. **d,** U2OS cell strains were plated at cloning densities into complete medium with increasing doses of calicheamicin. Colonies were stained after eight days, and percent survival was calculated. Error bars represent the standard error of the means for six independent experiments. **e,** Cell survival of Halo-Ku70 and Halo-Ku80 cells treated for the indicated time with 1 μM Halo-PROTAC changing to fresh media without the Halo-PROTAC ligand.

Supplemental Figure 2.

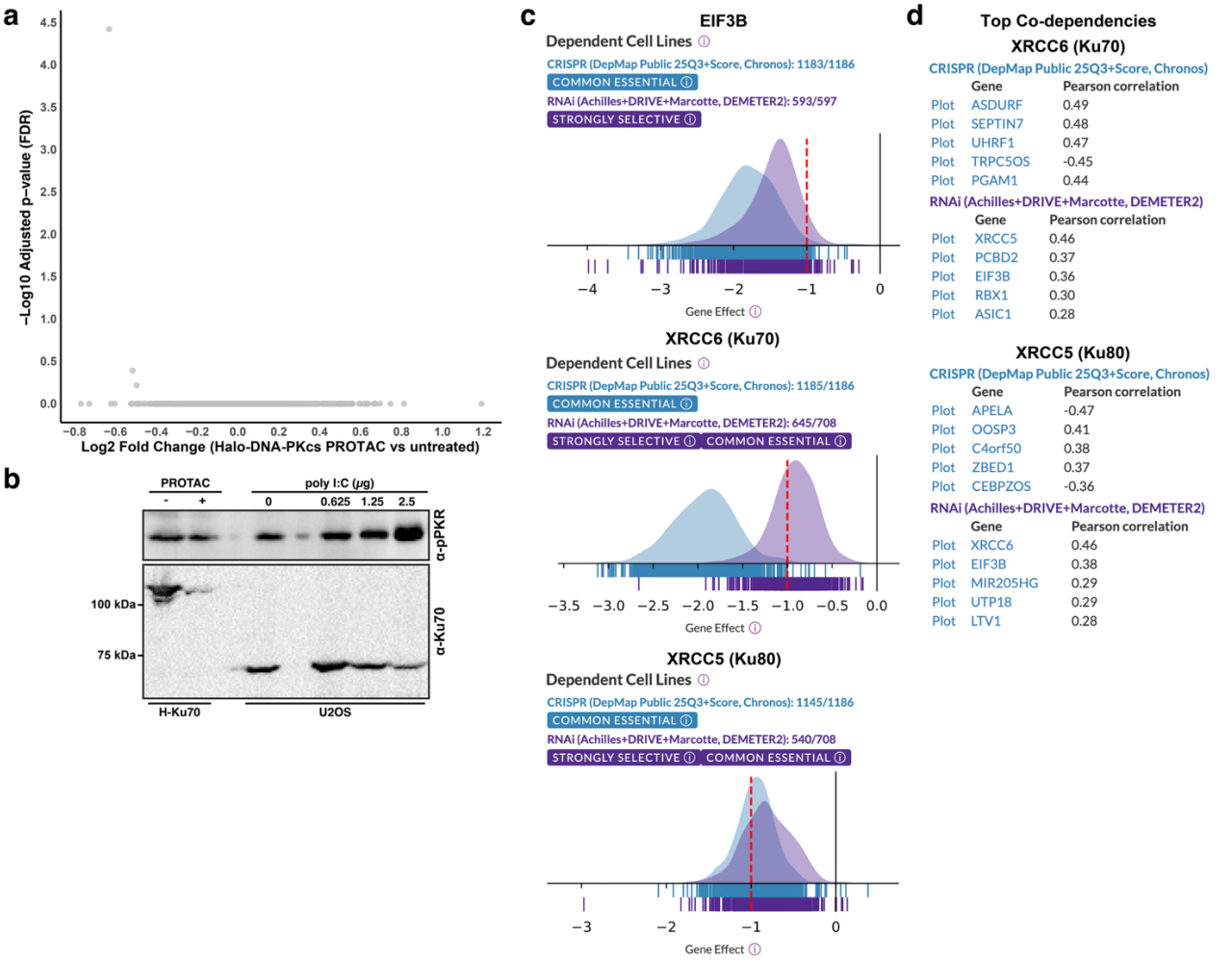


Figure S2. EIF3B is similarly cell essential as Ku70 and Ku80. a, Volcano plot comparing mRNA levels in untreated Halo-DNA-PKcs cells to cells treated with 1 uM Halo-PROTAC for 48 hours. **b,** Immunoblot demonstrating induction of T446 phosphorylation on PKR by transfection of poly I/C but not by 1 uM Halo-PROTAC treatment of Halo-Ku70 cells. **c,** The Dependency Map analyses of EIF3B, XRCC6 (Ku70), and XRCC5 (Ku80). All three are similarly essential in over 1000 cell lines as tested by CRISPR knockout experiments or in over 500 cell strains as tested by RNAi experiments. **d,** Top Dependency Map co-dependencies of Ku70 and Ku80.

Supplemental Figure 3.

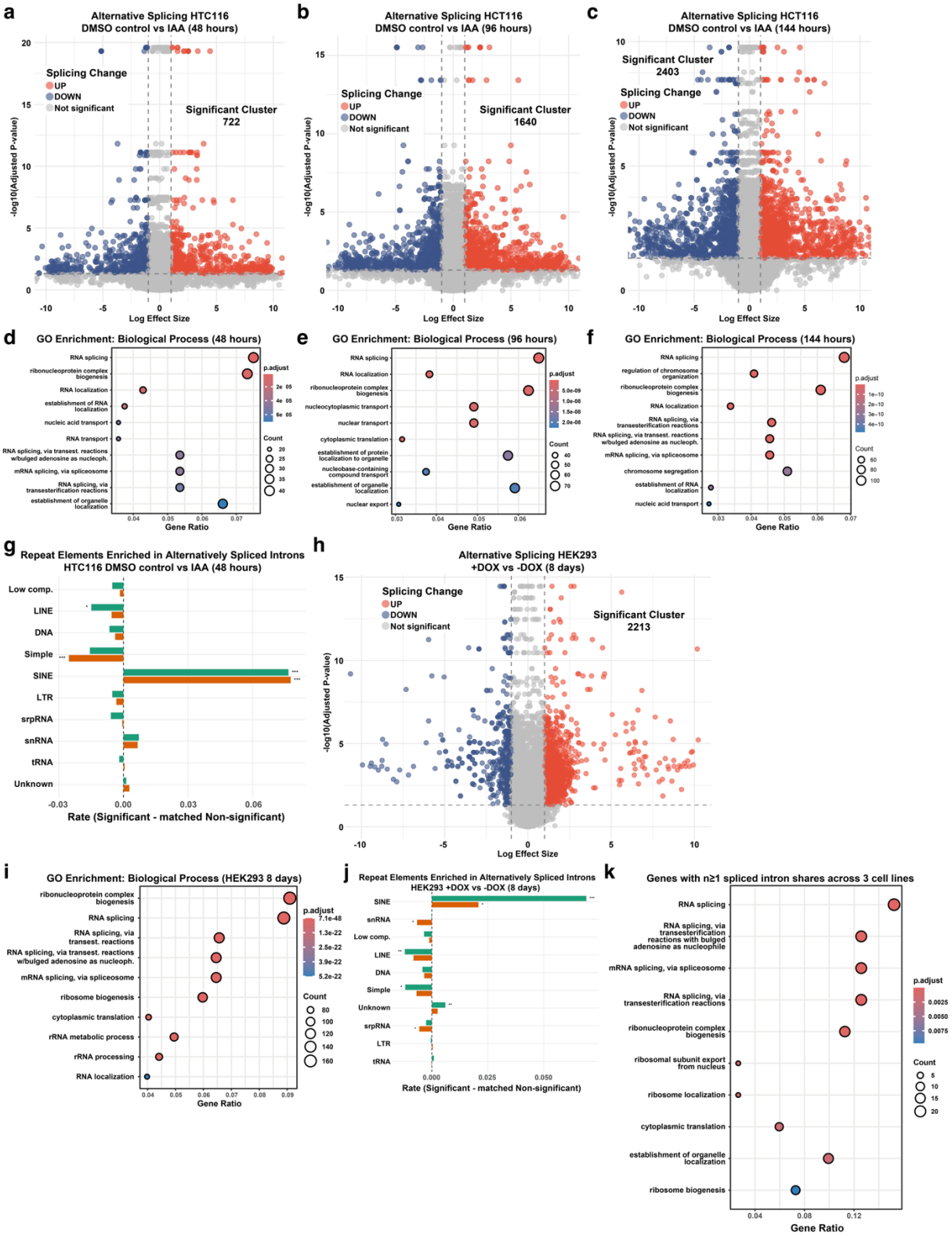


Figure S3. Depletion of Ku causes splicing defects in HEK293 and HCT116 cell lines. a-c, Volcano plot showing differential splicing events identified by LeafCutter in Ku80-depleted HCT116 cells at increasing timepoints. Each point represents an intron cluster, with significantly up- and downregulated events (|logef| > 1, adjusted P < 0.05) shown in red and blue, respectively; non-significant events are in grey. Dashed lines indicate effect size and significance thresholds. **d-f,** Gene ontology analysis of biological processes for genes with alternatively spliced introns in Ku80-depleted HCT116 cells at increasing time points. **g,** Analysis of the enrichment of repetitive elements enriched in alternatively spliced introns in HCT116 cells after Ku80 depletion for 48 hours. **h,** Volcano plot showing differential splicing events identified by LeafCutter in Ku70-depleted HEK293 cells. Each point represents an intron cluster, with significantly up- and downregulated events (|logef| > 1, adjusted P < 0.05) shown in red and blue, respectively; non-significant events are in grey. Dashed lines indicate effect size and significance thresholds. **i,** Gene ontology analysis of biological processes for genes with alternatively spliced introns in Ku70-depleted HEK293 cells. **j,** Analysis of the enrichment of repetitive elements enriched in alternatively spliced introns in HEK293 cells after Ku70 depletion for 48 hours. **k,** Gene ontology analysis of biological processes for genes with alternatively spliced introns in all three cells lines after Ku70 or Ku80 depletion.

Supplemental Figure 4.

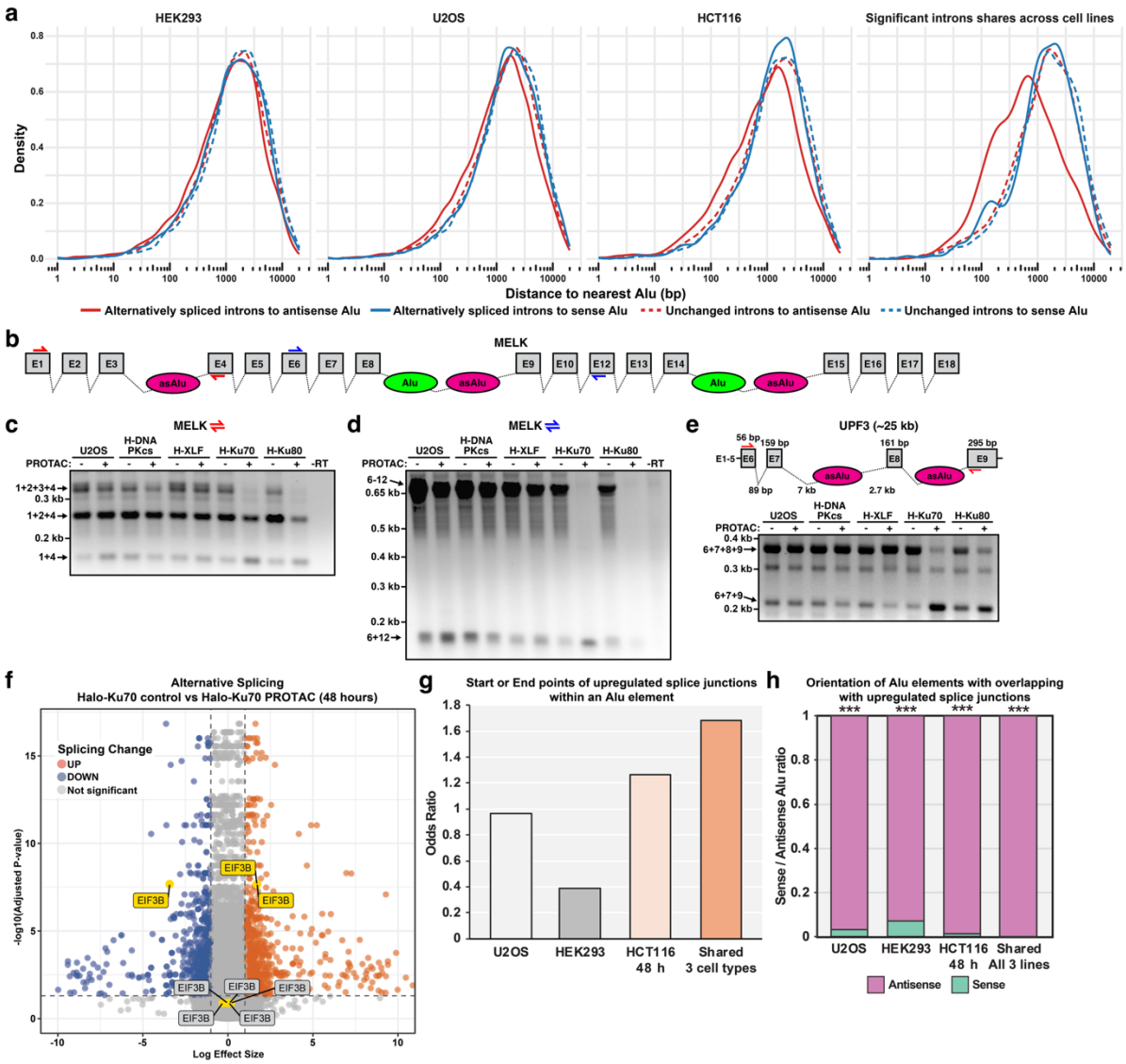


Figure S4. Splicing of antisense Alu elements is altered after Ku depletion. a, Density plots showing the genome-wide distance ($\log_{10}$ scale, in bp) from intron midpoints to the nearest Alu element located in sense (blue) or antisense (red) orientation relative to the host gene strand, for introns significantly (solid) or non-significantly (dashed) spliced (FDR < 0.05) in HEK293, U2OS, and HCT116 cells, and for introns significant across all three datasets ("Significant introns shared across cell lines"). **b,** Cartoon depicting the region of MELK transcript that undergoes differential splicing after Ku70 or Ku80 depletion. **c-d,** RT PCR using the indicated primers to assess splicing of *MELK* between exons 1-4 (**e**) and exons 6-12 (**f**). **e,** Cartoon and RT-PCR across the first four exons of the UPF3 mRNA in cell lines expressing HaloTagged NHEJ factors treated with Halo-PROTAC ligand. **f,** Volcano plot showing differential splicing events identified by LeafCutter in U2OS cells after Ku70 depletion, indicating positions of the EIF3B mRNA splice junctions. **g,** Odds ratios of splice junctions upregulated after Ku depletion containing splice donors or acceptors that overlap with Alu elements in U2OS, HEK293, and HCT116 cells, or for introns that are alternatively spliced in all three cell lines after Ku depletion. **h,** Orientation bias of Alu elements overlapping Ku-sensitive splice junctions. Non-canonical ΔPSI > 0 splice-junction endpoints that intersect Alu elements were classified as sense or antisense relative to host-gene transcription. Orientation counts were summarized for each dataset and for the reproducible 3-way set and tested against a 50:50 expectation using a binomial test.

Supplemental Figure 5.

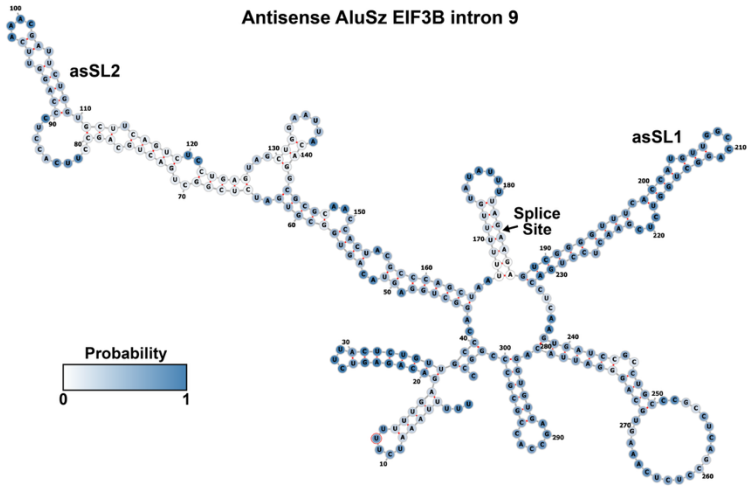
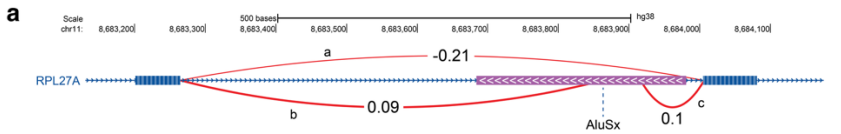
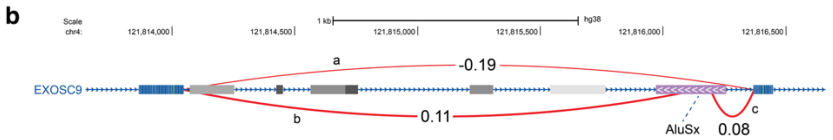


Figure S5. Secondary structure prediction of intron 9 of the EIF3B transcript, which contains a cryptic splice acceptor upstream of the asSL1 stem loop which is upregulated after Ku depletion in U2OS, HEK293, and HCT116 cells.

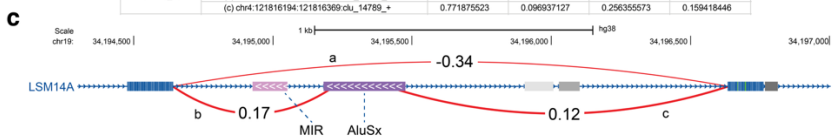
Supplemental Figure 6.



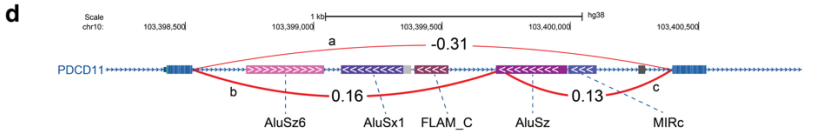
Cell line	intron	logef	control	Ku_depletion	deltapsi
U2OS	(a) chr11:8683265:8684006:clu_18859_+	-3.635759853	0.996459742	0.783889543	-0.212570199
	(b) chr11:8683265:8683863:clu_18859_+	0.68799938	0.001489459	0.068431798	0.086942339
	(c) chr11:8683917:8684006:clu_18859_+	0.709160718	0.001733642	0.105130727	0.103397084
HEK293	(a) chr11:8683265:8684006:clu_18859_+	-5.032624759	0.998908022	0.861887478	-0.137020544
	(b) chr11:8683265:8683863:clu_18859_+	-0.89173718	0.000955991	0.05184939	0.050893399
	(c) chr11:8683917:8684006:clu_18859_+	1.389383428	0.000134794	0.071554781	0.071419987
HCT116_48h	(a) chr11:8683265:8684006:clu_18859_+	-3.704813746	0.994659868	0.823530164	-0.171129704
	(b) chr11:8683265:8683863:clu_18859_+	-0.33808392	0.002727057	0.065441215	0.062714158
	(c) chr11:8683917:8684006:clu_18859_+	0.169208656	0.002608714	0.103967572	0.101358858



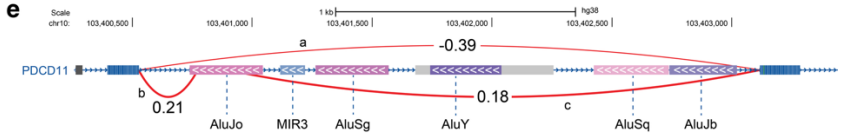
Cell line	intron	logef	control	Ku_depletion	deltapsi
U2OS	(a) chr4:121814047:121816369:clu_19621_+	-1.268549925	0.234024031	0.046824979	-0.187199052
	(b) chr4:121814047:121816144:clu_19621_+	0.570907688	0.410742473	0.517193481	0.106451008
	(c) chr4:121816194:121816369:clu_19621_+	0.559013843	0.341117249	0.42444525	0.083328001
HEK293	(a) chr4:121814047:121816369:clu_5074_+	-1.44399759	0.237152692	0.08372053	-0.153432163
	(b) chr4:121814047:121816144:clu_5074_+	0.927059167	0.038207071	0.144440466	0.106233389
	(c) chr4:121816194:121816369:clu_5074_+	0.714610399	0.061838468	0.189031496	0.127193027
HCT116_48h	(a) chr4:121814047:121816369:clu_14789_+	-2.073429974	0.169687827	0.026079714	-0.143608112
	(b) chr4:121814047:121816144:clu_14789_+	0.386327532	0.212423476	0.382044336	0.16962086
	(c) chr4:121816194:121816369:clu_14789_+	0.771875523	0.096937127	0.256355573	0.159418446



Cell line	intron	logef	control	Ku_depletion	deltapsi
U2OS	(a) chr19:34194641:34196634:clu_13265_+	-2.80788137	0.98220531	0.641409541	-0.340795769
	(b) chr19:34194641:34195219:clu_13265_+	0.269776729	0.012604267	0.178673706	0.16606944
	(c) chr19:34195413:34196634:clu_13265_+	2.262782221	0.001180145	0.122752291	0.121572147
HEK293	(a) chr19:34194641:34196634:clu_3595_+	-5.96618484	0.991269065	0.619076099	-0.372192967
	(b) chr19:34194641:34195219:clu_3595_+	-2.587850329	0.00873076	0.159880888	0.151150128
	(c) chr19:34195413:34196634:clu_3595_+	8.554035169	1.75E-07	0.221043014	0.221042839
HCT116_48h	(a) chr19:34194641:34196634:clu_10416_+	-2.293751752	0.287495204	0.205228265	-0.08226694
	(b) chr19:34194641:34195219:clu_10416_+	1.335407933	0.00168663	0.045368078	0.043681448
	(c) chr19:34195413:34196634:clu_10416_+	2.9350437	0.000137952	0.018372676	0.018234724



Cell line	intron	logef	control	Ku_depletion	deltapsi
U2OS	(a) chr10:103398528:103400397:clu_14933_+	-0.974291223	0.80534039	0.495819085	-0.309521305
	(b) chr10:103398528:103399734:clu_14933_+	0.584406902	0.085688299	0.250725944	0.165037645
	(c) chr10:103399830:103400397:clu_14933_+	0.694644951	0.059293093	0.193712496	0.134419403
HEK293	(a) chr10:103398528:103400397:clu_1162_+	-0.782950876	0.92420806	0.802665231	-0.121542829
	(b) chr10:103398528:103399734:clu_1162_+	0.569262242	0.018916966	0.063514792	0.044597826
	(c) chr10:103399830:103400397:clu_1162_+	0.213688635	0.056874974	0.133819977	0.076945003
HCT116_48h	(a) chr10:103398528:103400397:clu_2360_+	-1.069941648	0.588553981	0.467400318	-0.121153663
	(b) chr10:103398528:103399734:clu_2360_+	-0.429320874	0.064414575	0.097074445	0.03265987
	(c) chr10:103399830:103400397:clu_2360_+	-0.29567683	0.078528459	0.135266021	0.056737562



Cell line	intron	logef	control	Ku_depletion	deltapsi
U2OS	(a) chr10:103400528:103403118:clu_14934_+	-2.020567814	0.940024221	0.543925675	-0.396098545
	(b) chr10:103400528:103400772:clu_14934_+	0.132334127	0.053382963	0.265947939	0.212564975
	(c) chr10:103400897:103403118:clu_14934_+	1.888233687	0.006592816	0.190126386	0.18353357
HEK293	(a) chr10:103400528:103403118:clu_1163_+	-1.809039431	0.979312664	0.758772608	-0.220540055
	(b) chr10:103400528:103400772:clu_1163_+	0.500637509	0.012459123	0.097220469	0.084761346
	(c) chr10:103400897:103403118:clu_1163_+	1.308401922	0.008228213	0.144006923	0.13577871
HCT116_48h	(a) chr10:103400528:103403118:clu_2360_+	-1.182465641	0.246287577	0.174773933	-0.071513645
	(b) chr10:103400528:103400772:clu_2360_+	1.866769705	0.003204722	0.047983357	0.044778635
	(c) chr10:103400897:103403118:clu_2360_+	1.55879724	0.005184953	0.05705508	0.051870127

Figure S6. Additional examples of transcripts with cryptic splice sites in intronic antisense Alu elements. a-e, Diagram of splice junctions with start or end points in intronic antisense Alu elements, which are increased after Ku depletion in U2OS, HEK293, and HCT116 cells in the **(a)** *RPL27A,* **(b)** *EXOSC9,* **(c)** *LSM14A,* and **(d-e)** *PDCD11* genes.