Role of the SAF-A SAP domain in X inactivation, transcription, splicing, and cell proliferation

Judith A. Sharp[1], Emily Sparago[1], Rachael Thomas[1], Kaitlyn Alimenti[1], Wei Wang[1], and Michael D. Blower[1,2].

[1]Department of Biochemistry and Cell Biology

Chobanian and Avedisian School of Medicine, Boston University

72 E. Concord St, K112

Boston, MA 02118

[2] address correspondence to: mblower@bu.edu

Abstract

SAF-A is conserved throughout vertebrates and has emerged as an important factor regulating a multitude of nuclear functions, including lncRNA localization, gene expression, and splicing. SAF-A has several functional domains, including an N-terminal SAP domain that binds directly to DNA. Phosphorylation of SAP domain serines S14 and S26 are important for SAF-A localization and function during mitosis, however whether these serines are involved in interphase functions of SAF-A is not known. In this study we tested for the role of the SAP domain, and SAP domain serines S14 and S26 in X chromosome inactivation, protein dynamics, gene expression, splicing, and cell proliferation. Here we show that the SAP domain serines S14 and S26 are required to maintain XIST RNA localization and polycomb-dependent histone modifications on the inactive X chromosome in female cells. In addition, we present evidence that an Xi localization signal resides in the SAP domain. We found that that the SAP domain is not required to maintain gene expression and plays only a minor role in mRNA splicing. In contrast, the SAF-A SAP domain, in particular serines S14 and S26, are required for normal protein dynamics, and to maintain normal cell proliferation. We propose a model whereby dynamic phosphorylation of SAF-A serines S14 and S26 mediates rapid turnover of SAF-A interactions with DNA during interphase.

Introduction

Recent work has shown that the interaction of RNA with chromatin regulates the access of hundreds of different proteins to the DNA, thereby impacting chromatin structure and transcriptional control [1]. Interaction of RNA with chromatin can be controlled by proteins that contain both RNA-binding and DNA-binding domains. Prominent examples of RNA and DNA-binding proteins are the SAF-A and SAF-B proteins, which were originally described as factors that attach RNA to scaffold attachment regions (SARs) and factors that bind to hnRNA. [2-5].

SAF-A is a highly abundant nuclear protein that is essential in mice [6, 7] and cancer cells [8]. Mutations in SAF-A in humans or mice lead to a variety of neurological defects [9-11]. SAF-A has been implicated in regulation of mRNA splicing [6, 9, 10, 12], gene expression [9, 10], nuclear structure [13, 14], and tethering of the XIST and FIRRE lncRNAs to chromatin [15-18]. The SAF-A protein has several conserved functional domains, including an N-terminal SAP domain, a D/E-rich acidic region, a SPRY domain of unknown function, a central ATPase domain, and a C-terminal RGG domain involved in binding RNA and DNA (Figure 1a; [19]). At present, very little is known about which functional domains of SAF-A are required for each cellular process.

The SAP domain is a DNA binding domain present in many factors that regulate aspects of DNA metabolism [7]. The SAF-A SAP domain binds directly to SAR DNA *in vitro* [20, 21]. In addition, recent work has indicated that the SAF-A SAP domain can also bind directly to RNA both *in vitro* [22] and *in vivo* [23]. Two conserved serines present in the SAP domain, S14 and S26, are positioned in the DNA binding interface such that the electrostatic potential of the SAP domain is determined by the phosphorylation state of these two serines [15]. As a result, the increased negative charge associated with SAP domain S14 and S26 phosphorylation greatly reduces the affinity of SAF-A binding to DNA relative to the unphosphorylated protein. We previously demonstrated that SAP domain S14 and S26 phosphorylation by Aurora-B plays a critical role in regulating the interaction of SAF-A RNP complexes with the surface of mitotic chromosomes [15]. Since phosphopeptide evidence suggests that S14 and S26 are modified in interphase cells [24], we aimed to determine whether S14 and S26 phosphorylation regulates SAF-A nuclear functions.

In this study we present phenotypic analysis of cell lines expressing a SAP domain allelic series, including mutations which mimic the phosphorylated and unphosphorylated state, and a SAP domain deletion. We found that the SAP domain and phosphorylatable serines S14 and S26 were important for SAF-A function in XIST RNA localization and function, protein dynamics, and cellular proliferation. In contrast, the contribution of SAF-A to gene expression and mRNA splicing is largely independent of the SAP domain. Taken together, our data suggest dynamic phosphorylation of SAP domain serines S14 and S26 impacts a subset of SAF-A nuclear functions during interphase.

Results

The SAF-A SAP domain is required for cell proliferation

To investigate SAF-A nuclear functions in human cells, we first designed a cell line to degrade endogenous SAF-A in normal, diploid human RPE-1 cells immortalized with hTERT ([21]; Figure 1b). Both SAF-A genes were modified at the 3' end to include the minimal auxin-inducible degron sequence and a mCherry fluorescent tag (SAF-A-mAID-mCherry). A doxycycline-inducible TIR1 gene encoding the E3 ligase required for auxin-mediated protein degradation was integrated by lentivirus. Control experiments demonstrated efficient depletion of the SAF-A degron allele, resulting in complete knockdown of endogenous SAF-A in the vast majority of cells after 24 hours of adding doxycycline and 3-indole acetic acid (IAA; Figure 1d-e).

To evaluate the role of the SAP domain in SAF-A function, we then integrated doxycycline-inducible GFP-tagged rescue constructs encoding either full-length, wild-type SAF-A (SAF-A^{wt}-GFP), the phosphomutant S14A S26A (SAF-A^{AA}-GFP), the phosphomimetic allele S14D S26D (SAF-A^{DD}-GFP), or a deletion of the SAP domain (SAF-A$^{\Delta SAP}$-GFP). In this strategy, the addition of doxycycline and IAA triggers the simultaneous degradation of endogenous SAF-A and induction of GFP-tagged SAF-A rescue constructs. Analysis of SAF-A protein levels in these cell lines confirmed SAF-A-GFP expression and degradation of endogenous SAF-A-mAID-mCherry in cell extracts (Figure 1d lanes 4-7) and cell cultures (Figure 1e) under these conditions.

SAF-A is an essential gene required for cell proliferation and embryonic viability [6, 7]. We therefore tested whether our tagged SAF-A alleles supported essential functions of SAF-A by monitoring the number of cycling cells over 10 days, using PCNA and Ki-67 as markers for cell proliferation (Figure 1f-h). Both SAF-A-mAID-mCherry and SAF-A^{wt}-GFP maintained levels of PCNA- and Ki-67-positive cells comparable to wild-type RPE-1 cells, demonstrating that the tagged alleles encode functional proteins that complement the unmodified SAF-A gene and support essential functions of SAF-A. In contrast, SAF-A depletion caused a significant reduction in cycling cells over the time course of the experiment, as did the modification of SAP domain serines (SAF-A^{AA}-GFP and SAF-A^{DD}-GFP) and cells harboring a SAP domain deletion (SAF-A$^{\Delta SAP}$-GFP). Together, these data suggest that the SAP domain is required for essential functions of SAF-A, and that the two phosphorylatable serines S14 and S26 in the SAP domain are critical for the role of SAF-A in supporting normal levels of cell proliferation.

The SAF-A SAP domain serines are required for XIST RNA localization to the Xi

Several studies have reported that SAF-A is required for the proper localization of the noncoding XIST RNA to the inactive X chromosome (Xi) in female cells [16-18, 25, 26]. We note these studies used a variety of methods (siRNA, Cre-Lox recombination, and constitutive CRISPR SAF-A KO) that require several days to completely deplete SAF-A from cells, likely due to the high abundance of SAF-A protein. Since our SAF-A degron cell line resulted in robust depletion of SAF-A

within 24 hours of drug treatment (Figure 1), we determined whether the XIST RNA localization defect was present after this more acute period of depletion.

RPE-1 cells are female and harbor an Xi with XIST RNA expression and histone modifications associated with facultative heterochromatin [27]. XIST RNA is present at ~50 copies per nucleus in cell culture models of X inactivation [28-31], yet appears as a "cloud" in the confocal light microscope due to the resolution limit described by the point spread function. To evaluate the role of SAF-A in XIST RNA localization, we developed a quantitative imaging assay designed to count resolvable XIST RNA foci in 3D confocal optical stacks, reasoning that a localization defect would result in an increased number of XIST RNA molecules distal to the Xi, therefore also increasing the number of resolvable XIST RNA foci per nucleus. In this assay, cells were treated for 24 hours to deplete endogenous SAF-A-AID-mCherry and induce the SAF-A-GFP transgenes prior to XIST RNA FISH hybridization. 3D optical stacks were then acquired for $\geq$ 100 interphase nuclei per genotype in two biological replicate experiments. Images were analyzed in batch to count XIST RNA foci using a script we developed in Fiji software (Materials and methods). Representative XIST RNA FISH patterns are shown in Figure 2a; quantitative analysis of the two biological replicates is depicted in Figure 2b as a violin superplot [32].

Comparison of native RPE-1 cells (SAF-A^{wt}) with cells expressing SAF-A^{wt}-GFP demonstrated an equivalent number of XIST RNA foci per nucleus, demonstrating that the tagged construct complements for the role of SAF-A in XIST RNA localization to the Xi. In contrast, SAF-A depletion resulted in an average 2-fold increase in XIST RNA foci, with many events in the population having ~10-50 XIST RNA foci per cell (Figure 2b). These data confirm that SAF-A is required to maintain proper XIST RNA localization during this 24-hour experimental window, a period of time equivalent to one cell cycle in RPE-1 cells.

Analysis of cell lines with SAP domain mutations demonstrated that deletion of the SAP domain caused an increase in XIST RNA foci comparable to complete depletion of the SAF-A protein (Figure 2a-b). While the phosphomutant allele SAF-A^{AA}-GFP caused no detectable XIST RNA phenotype, the phosphomimetic allele SAF-A^{DD}-GFP resulted in potent XIST RNA mislocalization. These results indicate a critical role for the SAF-A SAP domain serines in maintaining proper XIST RNA localization to the Xi. While the allele modeling unphosphorylated SAP domain serines supports XIST RNA localization, the allele modeling constitutive phosphorylation of the two serines in the SAP domain antagonizes XIST RNA interactions with Xi chromatin.

The SAF-A SAP domain is required for maintenance of Polycomb-dependent histone modifications on the Xi

XIST RNA triggers the modification of Xi chromatin by recruiting the histone modification enzymes PRC1 to catalyze histone H2AK119ub, and PRC2 to catalyze histone H3K27me3, through a mechanism genetically linked to multiple XIST RNA effector proteins [33]. To investigate whether the SAF-A SAP domain is required to maintain XIST

RNA-dependent chromatin modifications on the Xi, we performed immunofluorescence for H2AK119ub and H3K27me3 and scored Xi enrichment in cell populations expressing wild-type and mutant SAF-A-GFP alleles, 24 hours after drug treatment. Comparison of native RPE-1 cell populations (SAF-A^{wt}) to those expressing SAF-A^{wt}-GFP demonstrated that both genotypes had similar frequencies of H2AK119ub and H3K27me3 Xi enrichment, again demonstrating the functionality of the GFP-tagged protein (Figure2c-d). In contrast, SAF-A depletion resulted in markedly decreased H2AK119ub and H3K27me3 Xi enrichment, with approximately half of the cell population having no detectable Xi enrichment of these histone modifications. This is consistent with previous results showing a loss of H3K27me3 on the Xi after siRNA-mediated depletion of SAF-A in mouse cells [16]. We also observed significantly reduced H2AK119ub and H3K27me3 Xi enrichment in cell lines expressing SAF-A^{AA}-GFP, SAF-A^{DD}-GFP, and SAF-$A^{\Delta SAP}$-GFP. Together these data demonstrate that the SAF-A SAP domain, and in particular SAP domain serines 14 and 26, are important for maintaining PRC1 and PRC2 functions on the Xi.

The SAF-A SAP domain contains an Xi localization signal

Immunofluorescent detection of SAF-A has led to conflicting reports about whether it is enriched or excluded from the Xi chromosome territory [34-36] suggesting that differences in sera or fixation conditions can result in contradictory observations of SAF-A localization. To avoid these potential problems, we performed live imaging of our SAF-A-GFP cell lines to determine whether SAP domain mutations resulted in altered Xi localization. We observed that SAF-A^{wt}-GFP was modestly enriched in the Xi chromosome territory as indicated by overlap with the Barr body, consistent with a prior report ([34], Figure 3a). SAF-A^{AA}-GFP and SAF-A^{DD}-GFP were also present in the Xi chromosome territory, with SAF-A^{AA}-GFP showing modest Xi enrichment more so than SAF-A^{DD}-GFP. In contrast, SAF-$A^{\Delta SAP}$-GFP was distinctly excluded from the Xi territory. These data show that there is an Xi localization sequence present in the SAF-A SAP domain that is largely independent of serines 14 and 26. It is likely that underrepresentation of SAF-$A^{\Delta SAP}$-GFP on the Xi is an important feature explaining the aberrant XIST RNA localization and Xi chromatin phenotypes associated with this mutation.

SAF-A exhibits rapid nuclear dynamics and interacts with nascent transcripts

To test whether the SAP domain plays a role in potentiating SAF-A nuclear dynamics in living cells, we performed FRAP analysis in cells expressing SAF-A-GFP transgenes. SAF-A^{wt}-GFP exhibited rapid nuclear dynamics such that the $t_{1/2}$ of recovery was an average 2.58 seconds, representing a high degree of nuclear mobility. These rapid nuclear movements suggest that most of the cellular pool of SAF-A participates in high turnover, transient engagements with chromatin or other nuclear structures. Indeed, we note that only 18% of SAF-A^{wt}-GFP was estimated to be immobile in a 30 second experiment (Figure 3b-d). Next, we compared SAF-A mobility in cells expressing the SAP domain allelic series:

SAF-A^{AA}-GFP, SAF-A^{DD}-GFP, and SAF-A$^{\Delta SAP}$-GFP. We found that all SAP domain mutations significantly decreased the $t_{1/2}$ of recovery to approximately 2.0 seconds, representing much more rapid nuclear motion than the wild-type protein. These data suggest that the SAP domain is engaging transiently with DNA/chromatin or other nuclear structures to impact the nuclear dynamics of SAF-A.

Recent work demonstrated that SAF-A localization and activity is responsive to nuclear transcription [1, 14], suggesting that SAF-A may interact with nascent transcripts. To determine whether transcription inhibition affects SAF-A mobility, we performed FRAP after inhibition of transcriptional elongation using the RNA polymerase II inhibitor LDC000067 (LDC) [37]. Interestingly, SAF-A^{wt}-GFP and all SAF-A SAP domain mutants exhibited increased nuclear mobility in the presence of LDC, suggesting that interaction of SAF-A with nascent transcripts is an important determinant of SAF-A nuclear dynamics. SAF-A$^{\Delta SAP}$-GFP had the strongest increase in mobility after transcription inhibition, with SAF-A^{AA}-GFP and SAF-A^{DD}-GFP having a more modest increase in mobility relative to the SAF-A^{wt}-GFP control. The SAF-A immobile fraction was reduced for all genotypes in the presence of LDC, suggesting there is a population of SAF-A molecules that are stably associated with nascent transcripts independently of the SAP domain (Figure 3d). Taken together, these data demonstrate that SAF-A is a highly dynamic protein that interacts with nascent transcripts and additional nuclear structures through the SAP domain.

Our FRAP data stands in contrast to a prior study which reported much slower rate of SAF-A nuclear motion [34]. We note there are experimental differences between that study and the data presented here. First, those experiments were conducted in transformed, aneuploid HEK293 cells which may have different nuclear properties than the system we describe in normal RPE-1 cells. In addition, the HEK293 cells were expressing SAF-A-GFP in addition to the endogenous protein, which could affect protein dynamics. We performed controls to ensure that SAF-A^{wt}-GFP is the sole source of SAF-A expression in the RPE-1 SAF-A degron cell line, and that the expression level was comparable to untagged native protein levels (Materials and methods). In addition, we have tested our SAF-A^{wt}-GFP reagent in multiple genetic assays and found that it complements for normal cell proliferation, Xi chromatin structure, global RNA transcription, RNA splicing, and cell cycle dependent localization (this study, [21]). Finally, our measured SAF-A dynamics are concordant with those reported for SAF-A expressed from a transgene in RPE-1 cells in a recent preprint [38]. For these reasons, we conclude that the rapid nuclear movements we observe with the SAF-A^{wt}-GFP allele are an accurate measurement of normal SAF-A dynamics.

Maintenance of Xi gene expression and Xi chromatin structure in SAF-A depleted cells

RPE-1 cells are a differentiated somatic cell type and therefore have an Xi chromosome in the epigenetically stable maintenance phase of X inactivation. During this phase, transcriptional silencing of the Xi can occur even if XIST RNA function is absent, and instead relies on other mechanisms such as DNA methylation to repress transcription [39-41].

We therefore tested whether the role for SAF-A on the Xi was solely epistatic to XIST RNA by monitoring for changes in Xi transcription and chromatin structure.

Recently, the haplotype-resolved genome sequence of human RPE-1 cells was published [42], suggesting that we could examine X-linked gene expression and chromatin structure using high-throughput sequencing techniques. To measure allele-specific expression we used the Personal Allele Caller software which was shown to have the highest accuracy and least mapping bias [43]. Examining RNA-seq data from wild-type RPE-1 cells revealed that the vast majority of autosomal genes have equal expression from the maternal and paternal alleles (Fig 4a). In contrast, genes located on the X-chromosome exhibit a strong bias with most expression arising from the 'a' allele. We note RPE-1 cells contain a 73 MB duplication of chromosome 10 translocated to the X chromosome [42, 44]. We detected a modest bias in allele-specific expression on chromosome 10 in our RNA sequencing data, suggesting that this translocation occurs onto the Xi. In contrast to most X-linked genes, 2 of 4 genes reported to escape X inactivation exhibited equal expression between both alleles (Figure 4a red points). Although there are likely more genes that escape Xi silencing, we could detect only 4 with SNPs in RPE-1 cells.

To further validate our ability to detect allele-specific differential expression of genes on the X chromosome, we performed ATAC-seq [45] and Cut-and-Run [46] for the histone modifications H3K4me3 and H3K27Ac. Consistent with our allele-specific RNA-seq analysis, we found that chromatin accessibility and active chromatin modifications are equally represented on both alleles on all autosomes except chromosome 10, again likely due to the duplication/translocation to the Xi (Fig 4c). In contrast, for the X chromosome we observed a decrease in chromatin accessibility on the 'b' allele (Fig 4c) and a concomitant decrease in active histone marks on the 'b' allele (Fig 4e, S1g). Taken together, the strong allelic bias in gene expression, chromatin accessibility, and active histone marks solely on the X chromosome strongly suggests that allele-specific sequencing is a viable strategy to monitor gene expression from the Xi in human RPE-1 cells.

To determine if SAF-A depletion resulted in reactivation of genes on the Xi, we used edgeR to compare gene expression from the 'a' (active) and 'b' (silenced) alleles in RPE-1 cells with and without SAF-A depletion. We found that genes located on the X chromosome exhibited nearly identical gene expression after 24 hours of SAF-A depletion (Figure 4b). To determine if SAF-A depletion over a longer time frame would reactivate gene expression from the Xi, we performed RNA-seq at 48 and 72 hours after SAF-A depletion. SAF-A depletion did not globally reactivate Xi gene expression under any of these conditions (Figure S1a-b). Similarly, none of the SAF-A SAP domain mutants reactivated gene expression on the Xi (Figure S1c-f). We also tested whether SAF-A depletion resulted in altered chromatin accessibility or active chromatin modifications on the Xi, using ATAC-seq and Cut-and-Run for the H3K4Me3 and H3K27Ac histone modifications. Consistent with our RNA sequencing results, we did not observe changes in the chromatin structure of the Xi following SAF-A depletion (Figure 4d, 4f and S1h).

Collectively, the data argue that SAF-A impacts XIST RNA localization and XIST-dependent organization of PRC1 and PRC2 activity on the Xi, but does not affect other parallel pathways enforcing transcriptional silencing in the maintenance phase of X inactivation. This is in contrast to the documented role of SAF-A in establishment of the Xi during cell differentiation [18]. We conclude that the role of SAF-A on the Xi is epistatic to XIST RNA.

SAF-A depletion results in transcriptional changes that accumulate over multiple cell divisions

To understand how the loss of SAF-A affects gene expression over time in RPE-1 cells, we performed RNA-seq on cells depleted for SAF-A for 24, 48, and 72 hours. Interestingly, we observed a steady increase in the number of misregulated genes during the time course of SAF-A depletion, with a modest number of misregulated genes at 24 hours (227), additional genes misregulated at 48 hours (313) and increasingly more genes misregulated at 72 hours (1196; Figure 5a-c). At each time point, the majority of misexpressed genes in SAF-A depleted cells exhibited reduced expression relative to control datasets Figure 5a-c). Gene ontology analysis of SAF-A downregulated gene categories over the time course identified a strong enrichment of terms associated with the extracellular matrix, cell adhesion, axon extension, and angiogenesis (Figure 5d). We note the modest effect of SAF-A on gene expression after acute depletion is consistent with a prior report [14]. These time-resolved datasets reveal that SAF-A depletion over multiple cell divisions causes cumulative changes in global gene transcription.

We tested whether the SAP domain is involved in the role of SAF-A maintaining proper gene expression (Figure 5e, S2). We performed RNA-seq and compared gene expression in cells expressing SAF-A^{wt}-GFP, SAF-A^{AA}-GFP, SAF-A^{DD}-GFP, and SAF-A$^{\Delta SAP}$-GFP, and SAF-A depleted cells. Figure 5e shows the cumulative distribution of gene expression for all genotypes. Surprisingly, we discovered that the gene expression profile of cells expressing SAP domain mutations was largely overlapping with cells expressing the wild-type allele. Only cells lacking SAF-A expression showed a significant deviation from the wild-type gene expression state (Supplemental Figure 2). We conclude that SAF-A maintains the transcriptome independently of the SAP domain.

The SAF-A SAP domain plays a minor role in mRNA splicing

Foundational work found that SAF-A regulates mRNA splicing in multiple organisms and tissue types [6, 12, 47]. Presently, it is not known whether splicing defects in SAF-A depleted cells are emergent at early time points after depletion, or whether the SAP domain is required to promote normal mRNA splicing. To understand how SAF-A contributes to mRNA splicing, we used rMATS to analyze differential splicing in SAF-A depleted cells and all SAP domain mutant cell lines at 24 hours from the mRNA-seq data described above. Consistent with previous work, SAF-A depletion

led to changes in the inclusion of ~1600 exons (FDR < 0.01; Figure 6a, 6c, S3a-b). SAF-A depletion tended to result in a loss of exon inclusion, but SAF-A was also required to promote exon exclusion for some transcripts. Gene ontology analysis of overrepresented functional groups exhibiting altered splicing revealed that SAF-A is required for the splicing of mRNAs encoding proteins important for transiting the nuclear pore, cell cycle control, and microtubule organization (Figure 6d). Comparison of changes in mRNA splicing to changes in gene expression revealed no correlation, suggesting that changes in mRNAs splicing do not directly cause changes in gene expression (Figure S3c).

To test whether the SAF-A SAP domain was involved in mRNA splicing, we compared splicing changes in cells expressing SAF-A^{wt}-GFP, SAF-A^{AA}-GFP, SAF-A^{DD}-GFP, and SAF-$A^{\Delta SAP}$-GFP, and SAF-A depleted cells Figure 6b). Comparison of these genotypes revealed that SAF-A depleted cells showed the most significant shift from the wild-type condition (p-value 2.2×10^{-16} SAF-A^{wt}-GFP vs SAF-A depletion), with more minor changes observed in cells expressing SAF-$A^{\Delta SAP}$-GFP (p-value 1.375×10^{-8} SAF-A^{wt}-GFP vs SAF-$A^{\Delta SAP}$-GFP, Figure 6c). In contrast, mutation of the SAP domain serines in cells expressing SAF-A^{AA}-GFP and SAF-A^{DD}-GFP had very little effect on splicing.

To confirm the observed mRNA splicing changes we used RT-PCR with primer pairs spanning regulated exons. We tested 4 mRNAs exhibiting both increased and decreased exon inclusion in the SAF-A mutants and confirmed the magnitude and direction of all changes predicted by mRNA sequencing (Figure S4). We conclude that SAF-A is important for mRNA splicing and that that the SAP domain plays a minor role in this process.

To understand whether splicing changes accumulate over time after SAF-A depletion, we analyzed splicing changes at 24, 48 and 72 hours. Surprisingly, we did not observe a phenotypic gradient over time with mRNA splicing as we had observed with gene expression. Splicing changes were immediately apparent 24 hours after depletion, and the severity of splicing defects did not increase over time (Fig S5). We conclude that acute depletion of SAF-A leads to dramatic changes in splicing, suggesting that SAF-A plays a direct role in this process.

Splicing can be regulated by protein binding to pre-mRNA transcripts at splicing enhancer and repressor sites both within exons, and upstream and downstream of regulated exons [48]. To determine if specific sequence motifs are present in or around SAF-A regulated exons we used MEME [49] and a custom k-mer enrichment program [50] to analyze various regions in and around SAF-A regulated exons. We did not detect significant enrichment of any sequence motifs in any class of sequence (Figure S6a-d). To determine if SAF-A binding is enriched in or around SAF-A regulated exons we utilized SAF-A eCLIP data from the ENCODE consortium from two different cell lines. We did not detect any SAF-A enrichment upstream, downstream or in regulated exons in the eCLIP data, consistent with a wide distribution of SAF-A binding throughout the transcriptome (Figure S7a) [51]. In addition to sequence motifs, it is possible SAF-A could recognize secondary structure in mRNA. We used mfold [52] to predict the free energy of control and SAF-A regulated exon regions, but detected no differences related to SAF-A regulation. We also found that SAF-A splicing regulation was

not correlated with gene or intron length (Figure S7b-d). In contrast, when we compared SAF-A regulated transcripts to RNAs identified as being stably associated with SAF-A [21], we found that transcripts containing exons that exhibited increased inclusion after SAF-A depletion were significantly enriched in SAF-A IP samples (p-value 4.9×10^{-9}, Figure 6e). This suggests that stable association of SAF-A with a transcript may promote exon exclusion, but that the mechanism targeting SAF-A to specific transcripts in currently unclear.

Discussion

In this study, we describe how a degron system in normal, diploid RPE-1 cells can be used to explore SAF-A function in a number of genetic assays. Given the high abundance of SAF-A in the nucleus [53], one advantage of this system is that it allows for the rapid depletion of SAF-A, allowing for observation within 24 hours, as well as longer time periods. This allowed us to obtain time resolved datasets for the role of SAF-A in cell proliferation, gene expression, and splicing. The stable integration of inducible SAF-A transgenes allowed for precise titration of SAF-A expression level, so as to approximate normal protein levels as closely as possible. In addition, because the SNPs present in RPE-1 cells have been mapped to each haplotype genome, this allowed us to examine X-linked gene expression with allele-specific resolution. This RPE-1 degron system engineered to simultaneously deplete endogenous SAF-A and induce tagged SAF-A transgenes therefore represents a set of valuable reagents to study SAF-A cellular function.

We used this system to explore the role of the SAF-A SAP domain and SAP domain phosphorylation in supporting XIST RNA localization. In addition, we developed a quantitative method to evaluate XIST RNA localization among several genotypes (Materials and Methods). Using this assay, we found a SAP domain deletion is defective for XIST RNA localization to the Xi in human cells, consistent with a prior study in mouse cells [16]. Interestingly, the SAF-A alleles modeling the unphosphorylated (SAF-A^{AA}-GFP) and phosphorylated state (SAF-A^{DD}-GFP) had opposite XIST RNA localization phenotypes: SAF-A^{AA}-GFP had normal XIST RNA localization, and SAF-A^{DD}-GFP showed a disruption in XIST RNA localization. These phenotypes are likely attributable to differences in the affinity of the SAP domain for DNA [15]. While the SAF-A^{AA}-GFP allele could allow for XIST RNA attachment to Xi chromatin, the reduced DNA affinity present in SAF-A^{DD}-GFP would antagonize such interactions.

Interestingly, both SAF-A^{AA}-GFP and SAF-A^{DD}-GFP were defective in maintaining PRC1- and PRC2-dependent histone modifications on the Xi. XIST RNA is required to organize PRC1 and PRC2 activity on the Xi, but multiple studies suggest this occurs in a manner requiring one or more adapter proteins to bridge the interaction, such as hnRNP K or others [33]. Taken together, these data argue that the dynamic phosphorylation of SAF-A on serines S14 and S26 is required for formation of a functional, mature XIST RNP. Since SAF-A localizes along the length of XIST RNA [51], and SAF-A interacts with nascent transcripts (Figure 3 and [15]), it is possible that SAF-A is interacting with newly transcribed XIST RNA molecules, facilitating contacts with effector proteins along the length of the RNA.

We examined SAF-A nuclear dynamics and found that most of the SAF-A population is engaged with high turnover interactions with chromatin or other nuclear structures. The observation that SAF-A is a highly dynamic protein prompts a re-evaluation of two concepts in the field. First, the data argue that most SAF-A molecules are not part of a static nuclear structure [54]. Second, the data suggest that the SAF-A RNP makes rapid, transient engagements with chromatin, rather than being a stable tether to bridge RNA molecules to chromatin. Therefore, it is likely that SAF-A

molecules bound to XIST RNA also make high turnover contacts with Xi chromatin, driven by dynamic phosphorylation on SAP domain serines S14 and 26.

There are conflicting observations in the literature about whether SAF-A is enriched or depleted on the Xi in fixed cells [36, 55]. We observed the modest enrichment of SAF-A^{wt}-GFP on the Xi only in live imaging experiments, suggesting that there are changes occurring during cell permeabilization or fixation that are not accurately reporting the localization of SAF-A to the Xi. Therefore, we urge caution when interpreting SAF-A localization in fixed cells. Surprisingly, cells expressing the SAF-A$^{\Delta SAP}$-GFP allele showed marked exclusion from the Xi territory labeled by the Barr Body. These data suggest there is an Xi-localization signal within the SAP domain that is largely independent of SAP domain phosphorylation.

Previous work demonstrated that short-term depletion of SAF-A using siRNAs resulted in very few changes in gene expression [14]. Other studies in mice and human iPS cells found that depletion or mutation of SAF-A results in changes in the expression of hundreds of genes [9-11]. In this study, we used our system to explore the kinetics of gene expression changes in cells lacking SAF-A function and found that gene expression changes accumulate over multiple cell cycles, in a manner independent of the SAP domain or SAP domain phosphorylation. It is possible that this occurs through an indirect mechanism, such as a misexpressed gene or misspliced RNA causing long term amplification of gene expression changes. It is also possible that SAF-A affects a low turnover element of chromatin, such as a long-lived histone modification or chromatin modifier. Future work is needed to understand how SAF-A maintains the gene expression state, and which other conserved domains contribute to this function.

In contrast to gene expression, the impact of SAF-A depletion on splicing, XIST RNA localization, and XIST RNA-dependent histone modifications was observed within the first cell cycle of depletion, suggesting that SAF-A is directly linked to each of these functions, and is continually required to execute each of these functions. Importantly, the ability to compare alleles across assays in this study revealed that the phosphomimetic allele, SAF-A^{DD}-GFP, represents a separation of function mutation: whereas SAF-A^{DD}-GFP has a normal profile for mRNA splicing, SAF-A^{DD}-GFP failed to enforce normal XIST RNA localization on the Xi. Therefore, the role of SAF-A in maintaining XIST RNA localization is independent of mRNA splicing. In addition, the SAF-A^{AA}-GFP could execute normal XIST RNA localization but was defective in the recruitment of PRC1 and PRC2-dependent histone modifications. We note that localization and silencing activity are separable functions of the XIST RNA and are attributable to different functional repeat sequences on the RNA [56].

Our studies of cell cycle markers show that the SAF-A SAP domain, and SAP domain serines S14 and S26, are required to support normal cell proliferation. These data suggests that dynamic phosphorylation of S14 and S26 is important to fulfill essential roles of SAF-A. It is likely that this role is critical to maintaining rapid cell divisions during early

development and could explain why knockout of the mouse Saf-a gene results in early embryonic lethality [9]. In addition, we observed misexpressed genes in a number of categories important for normal organismal function (Figure 5; [8]). We conclude the constellation of phenotypes we present here provide insight for SAF-A SAP domain functions at both the cellular and organismal levels.

Materials and Methods

Cell culture and drug treatment

hTERT-immortalized RPE-1 cells were a gift from Brian Chadwick (Florida State University, Tallahassee, FL; ATCC CRL-4000) and were cultured as described [21]. The SAF-A degron cell line in RPE-1 cells has been described [21]. Lentiviruses with SAF-A^{wt}-GFP, SAF-A^{AA}-GFP, SAF-A^{DD}-GFP, and SAF-A$^{\Delta SAP}$GFP were prepared and used to transduce the SAF-A-degron cell line. Individual clones were expanded and validated by immunofluorescent detection of GFP and western blot analysis.

To induce SAF-A depletion +/- SAF-A-GFP expression, cells were treated with 0.5 μg/mL doxycycline (D3447; MilliporeSigma) and 100 μg/mL 3-indole acetic acid (IAA, I-5148 MilliporeSigma) for 24 hours. Control experiments determined that treatment with 0.5 μg/mL doxycycline resulted in SAF-A-GFP expression levels equivalent to the native, untagged protein (unpublished data). For the long-term depletion of SAF-A, media was replaced with fresh drug every 2-3 days.

Plasmids

GFP-tagged SAF-A alleles were cloned into the lentiviral expression vector pLVX-TetOne-puro (Takara Bio). The lentiviral plasmids pMB1103 (SAF-A^{wt}-GFP), pMB1109 (SAF-A^{AA}-GFP), and pMB1311 (SAF-A A^{DD}-GFP) have been described [15]. pMB1316 (SAF-A$^{\Delta SAP}$GFP) was prepared similarly.

Antibodies

Primary antibodies used in this study were as follows: mouse anti-GFP (sc-9996, Santa Cruz Biotechnology), rabbit anti-mCherry (600-401-p16, Rockland), mouse anti-tubulin (DM1A, MilliporeSigma), GFP booster nanobody-ATTO488 (GBA488, Bulldog Bio), RFP booster nanobody-ATTO 594 (RBA594, Bulldog Bio), mouse-anti histones H1-H4 (MAB3422, MilliporeSigma), rabbit anti Ki-67 (ab16667, Abcam), rabbit anti-H2AK119ub (D27C4, Cell Signaling Technology), mouse anti-H3K27me3 (39537, Active Motif), and human anti-PCNA (AK, a gift from Dr. Yoshinari Takasaki (Juntendo University School of Medicine, Japan).

Secondary antibodies used in western blotting were: donkey anti-mouse conjugated to 680RD (925-68072, LiCor) and donkey anti-rabbit conjugated to 680RD (925-68073, LiCor). Secondary antibodies used in immunofluorescent studies were donkey anti-mouse Alexa Fluor 647 (715-605-150, Jackson Immunoresearch), donkey anti-human Alexa Fluor 488 (709-545-149, Jackson Immunoresearch), donkey anti-human Cy3 (709-165-149, Jackson Immunoresearch), donkey anti-rabbit Alexa Fluor 488 (711-545-152, Jackson Immunoresearch), donkey anti-rabbit Cy3 (711-165-152, Jackson Immunoresearch), donkey anti-rabbit Alexa Fluor 647 (Jackson Immunoresearch), and donkey anti-mouse Cy3 (715-165-150, Jackson Immunoresearch.

Image acquisition and analysis

All images were acquired using a Nikon A1R confocal microscope equipped with a 60 × 1.4NA Vc lens and laser lines at 405, 488, 562, and 647 nm. Images were collected using Nikon Elements software driving the galvano scanner. Images were acquired as *Z* stacks spaced 0.2 μm apart. Each experiment was repeated in at least two biological replicates. For immunofluorescent detection of SAF-A alleles and Ki-67 (Figure 1e and 1g), we used the same fixation and detection conditions we previously described [21]. For PCNA and immunofluorescent detection of histone H3K27me3 and histone H2AK119ub, we performed permeabilization with CSK buffer (100 mM NaCl, 300 mM sucrose, 3 mM MgCl2, 10 mM PIPES pH 6.8) for 30 seconds, followed by treatment with CSK buffer + 0.5% Triton-X-100, followed by CSK buffer for 30 seconds, and fixation with 4% paraformaldehyde (15710, Electron Microscopy Sciences) buffered with 1X PBS. Our FISH conditions to detect XIST RNA have been published previously [21, 57, 58]. Images were rendered in Fiji software as a projection or single optical slice, as specified in the figure legends.

Live imaging and FRAP

For live imaging of SAF-A-GFP alleles on the Xi, cells were plated onto fluorodishes (FD35-100, World Precision Instruments) and treated with doxycyline and IAA for 24 hours prior to observation. Hoechst 33342 dye (Molecular Probes) was added just prior to imaging. Cells were imaged on a Nikon A1R confocal microscope equipped with a stage-top incubator and CO_2 chamber (Tokai Hit). For FRAP analysis, cells were plated as above, except without Hoechst dye. FRAP was performed as follows: the scan area was set to 128 x 128, at a scan zoom size of 8X. 10 images were acquired prior to bleaching. Bleach area was a 3x3 square. Bleaching was accomplished with 80% power of the 488 laser for 1 second. Images were acquired every 65 ms for 30 seconds following bleaching (456 images). For each allele, FRAP movies were collected from at least 25 cells in total on two different days. For experiments involving the transcriptional inhibitor LDC, the drug was added to the culture media 30 minutes prior to imaging. Movies were imported into Fiji and quantitative analysis of FRAP data was performed as described [59]. FRAP curves were plotted in R from average values across all cells analyzed.

ATAC-seq

1 million RPE-1 or SAF-A KD cells were used as input for the Zymo-seq ATAC-seq kit (D5458). Libraries were constructed according the manufacturers instructions. Lubraries were sequenced to a depth of ~120 million reads at Novogene using paired-end 150bp reads.

Cut-and-Run

1 million RPE-1 or SAF-A depleted cells were used as input for the EpiCypher CUTANA ChIC/CUT&RUN Kit Version 3 (14-1048) with anti-H3K4me3 (EpiCypher 13-0060), anti-H3K27Ac (Invitrogen, MA5-23516) or IgG. 5 ng of CUT&RUN enriched DNA were used as input for library construction with EpiCypher CUT&RUN Library Prep Kit (14-1001). Libraries were sequenced at a depth of ~20 million reads at Novogene using paired-end 150bp reads.

RNA-seq

rRNA depleted samples were converted into libraries using the NEB Next Ultra II RNA library kit. Libraries were sequenced at Novogene using 150 bp paired-end reads.

RNA-seq analysis.

Common QC and processing steps

All libraries were sequenced at Novogene using paired-end 150bp reads. Raw fastq files were analyzed for quality, adaptor sequences were trimmed, and duplicate reads were removed using the fastp software suite [60]. Unique, trimmed fastq files were used for all subsequent steps.

Differential gene expression analysis

De-duplicated, trimmed reads were aligned to the UCSC hg38 human genome sequence using Rsubread/subread [61]. Reads per gene were quantitated using FeatureCounts in R with Entrez Gene IDs as the reference. Differential gene expression analysis between control and knockdown or mutant cell lines was performed using edgeR [62]. For the plots in the manuscript we considered genes significant if they exhibited a logFC of >= 1 and FDR value less than 0.01. All MD plots were created using ggplot2 in R.

Gene Ontology Analysis

Significantly altered genes were selected from edgeR analysis. Differential genes were split into up- and down-regulated genes and submitted to EnrichR [63] Gene Ontology analysis using Biological Process and Benjamani Hofberg p-value correction. Enriched categories were manually curated to combine redundant terms and plotted using the enrichR results and ClusterProfiler [64].

Alternative splicing analysis

De-duplicated, trimmed reads were used as input for rMATS [65] in conjunction with the ENSEMBL hg38 and ENSEMBL transcript annotations. RMATS output files were filtered as follows: first, we only considered splicing events supported by an average of 12.5 junction spanning reads per sample (e.g. for a comparison including two control and two experimental samples we required a total of 50 reads). For determination of the number of significant changes we filtered for a FDR < 0.01. To produce Sashimi plots we used rmats2sashimi.py. To conduct GeneOntology analysis of alternative splicing changes we used the ENSEMBL Gene names of significantly changed exons as input for EnrichR and ClusterProfiler. For analysis of conserved domains altered by alternative splicing we used SpliceTools and batch blast against the NCBI conserved domain database.

Motif analysis of alternatively spliced exons

To search for motifs near alternatively spliced exons we used bedtools to collect 100bp upstream and downstream of regulated exons and unregulated exons (as a control). We also used bedtools to collect the sequences of all regulated

exons and unregulated exons. We then used these sequences as input for STREME for discriminative identification mode. As an alternative approach we used these sequences as input for a custom Perl script that calculates the frequency of all k-mers in a set of experimental and control sequences then derived an enrichment score for each k-mer. For this analysis we analyzed hexamers but obtained similar results for k-mers ranging from 5-8. For secondary structure DeltaG prediction we used all the sequences described above as input for the Vienna RNAfold package. DeltaG values were extracted from the batch output using R and compared between regulated exons and nonregulated exons.

Allele-specific gene expression analysis

We used the Personalized Allele Caller [43] software in conjunction with the RPE-1 .vcf file [42] of maternal and paternal SNPs to generate counts per gene from each RNA-seq library. To validate the RPE-1 .vcf file with our own batch of RPE-1 cells we sequenced the genome of these cells to a depth of 20X using paired-end 150bp reads. We then ran PAC on this DNA sequencing data using the published .vcf file. We eliminated all SNPs with 0 counts for either allele in the DNA sequencing experiment and all SNPs that exhibited a allele-specific bias of > 1 Standard Deviation from the mean. We used the filtered .vcf file for all subsequent experiments. We filtered the genes to only consider genes with a total of greater than 10 counts per gene. To compare maternal:paternal ratios in different conditions we created average maternal:paternal ratios for each gene from pairs of control and SAF-A depleted/mutant cells. For comparison of RPE-1 to SAF-A depleted cells, this consisted of 6 libraries for each condition. For comparison of RPE-1 to each SAF-A mutant it consisted of 2 libraries for each condition. For the time series of SAF-A depletion, it consisted of 2 libraries for each condition. Averaged ratios were used to create violin plots in R.

DGE analysis of gene reactivation on the inactive X chromosome

To determine if genes are reactivated on the inactive X chromosome we used edgeR to calculate a logFC and FDR value for a:b ratios for each gene on the X chromosome. We then compared the logFC (a/b) and FDR values for different genotypes. The number of libraries compared for each condition is described above.

ATAC-seq library QC

Sequencing .fastq files were initially QCed using fastp as described above. Reads were aligned to hg38 using bwa-mem. Output .bam files were analyzed for fragment size and genomic location using ATACseqQC [66]. To perform allele-specific ATAC-seq QCed trimmer, deduplicated fastq files were used as input for PAC with the modified .vcf file described above.

CUT&RUN analysis

Sequencing fastq files were QCed using fastp as described above and aligned using PAC as described above. Allele specific read counts for IgG were subtracted from the read counts at the same gene in the antibody libraries to remove background. Genes were filtered for >= 0 reads on for each allele and >5 reads total. The ratio of “a” to “b” allele were calculated for each gene and averaged across replicates.

Quantitation of number of XIST particles in SAF-A depleted and SAP mutant cells

3D images of XIST localization in RPE-1, SAF-A depleted cells, and in all SAF-A-GFP cell lines were acquired in 0.2μm z stacks. To quantitate the number of XIST RNA particles in z-stack confocal images, we used a macro implemented in FIJI. Briefly, we used the DAPI signal to create a mask for nuclei in each image slice after performing a Gausian blur, Otsu threshold, fill holes and exclude nuclei touching the image edges. Nuclei were then added to the 3D ROI Manager to create linked 3D objects. We then detected XIST RNA particles by segmenting the RNA FISH image using the Otsu method. XIST particles were then added to the 3D ROI manager. We then used the 3D ROI Manager to measure properties of all particles, measure colocalization between all particles, and measure distances between all particles. Three measurement files were saved for each image. We then used a custom Perl script to summarize the measurements for each nucleus from these three files. Measurements were performed in a batch manner with no user supervision.

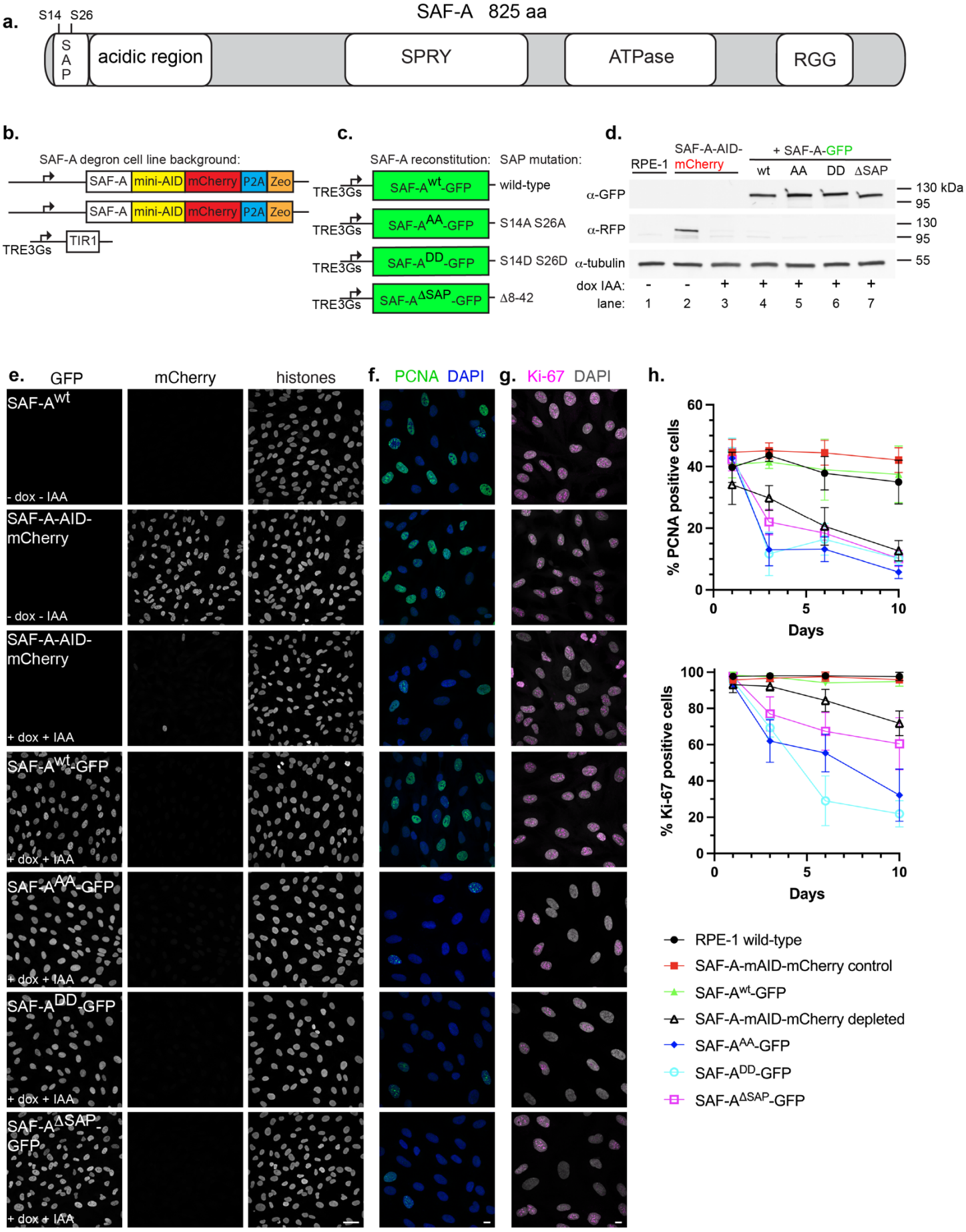

a.
S14 S26
SAF-A 825 aa
S A P
acidic region
SPRY
ATPase
RGG
b.
SAF-A degron cell line background:
SAF-A mini-AID mCherry P2A Zeo
SAF-A mini-AID mCherry P2A Zeo
TRE3Gs TIR1
c.
SAF-A reconstitution:
SAP mutation:
TRE3Gs SAF-A^wt-GFP wild-type
TRE3Gs SAF-A^AA-GFP S14A S26A
TRE3Gs SAF-A^DD-GFP S14D S26D
TRE3Gs SAF-A^ΔSAP-GFP Δ8-42
d.
RPE-1
SAF-A-AID-mCherry
+ SAF-A-GFP
wt AA DD ΔSAP
α-GFP
α-RFP
α-tubulin
130 kDa
95
130
95
55
dox IAA: - - + + + + +
lane: 1 2 3 4 5 6 7
e. GFP mCherry histones
SAF-A^wt
- dox - IAA
SAF-A-AID-mCherry
- dox - IAA
SAF-A-AID-mCherry
+ dox + IAA
SAF-A^wt-GFP
+ dox + IAA
SAF-A^AA-GFP
+ dox + IAA
SAF-A^DD-GFP
+ dox + IAA
SAF-A^ΔSAP-GFP
+ dox + IAA
f. PCNA DAPI
g. Ki-67 DAPI
h.
% PCNA positive cells
Days
% Ki-67 positive cells
Days
RPE-1 wild-type
SAF-A-mAID-mCherry control
SAF-A^wt-GFP
SAF-A-mAID-mCherry depleted
SAF-A^AA-GFP
SAF-A^DD-GFP
SAF-A^ΔSAP-GFP

Figure 1. The SAP domain and phosphorylatable SAP domain serines S14 and 26 are required for cell proliferation. A. Schematic depicting full-length SAF-A (isoform a, 825 amino acids) with domains drawn to scale. B. Design of the SAF-A-AID-mCherry degron to replace the endogenous SAF-A genes in RPE-1 cells. C. SAF-A transgenes with alanine (phosphomutant) or aspartic acid (phosphomimetic) mutations at positions S14 and S26, or expressing a SAP domain deletion. D. Western blot analysis of SAF-A cell lines to monitor expression of SAF-A-AID-mCherry (α-RFP panel) and SAF-A-GFP (α-GFP panel) in cell lines. Tubulin was used as a loading control. Lane 1: untagged RPE-1 parental cell line. Lane 2: SAF-A degron cell line, either without (lane 2), or with (lane 3) doxycycline and IAA. Lanes 4-7: cells expressing SAF-A-GFP transgenes, treated with doxycycline and IAA. Molecular weight markers are indicated to the right of the panel. E. Immunofluorescence of cells expressing tagged SAF-A, as indicated by the panel insets. Treatment with doxycycline and IAA as indicated led to complete depletion of SAF-A-AID-mCherry in the vast majority of cells. Integrated SAF-A-GFP transgenes showed uniform expression in the cell population after induction. Bar, 40 μm. F-G. PCNA and Ki-67 immunofluorescence to monitor cell proliferation. Cells were treated with doxycycline and IAA to induce endogenous SAF-A depletion and SAF-A-GFP transgene expression, and compared to control cell populations. Cells were treated for 1, 3, 6, or 10 days as indicated. SAF-A expression is the same as the panel insets shown in E. All images in panels E-G are rendered as a maximum projection of a 3D stack. Bar, 10 μm. H. Quantitation of cell populations with PCNA- and Ki-67-positive cells expressed as percent of the total population. 300 cells were scored for PCNA and Ki-67 immunofluorescence in two biological replicates; the average and standard deviation is shown.

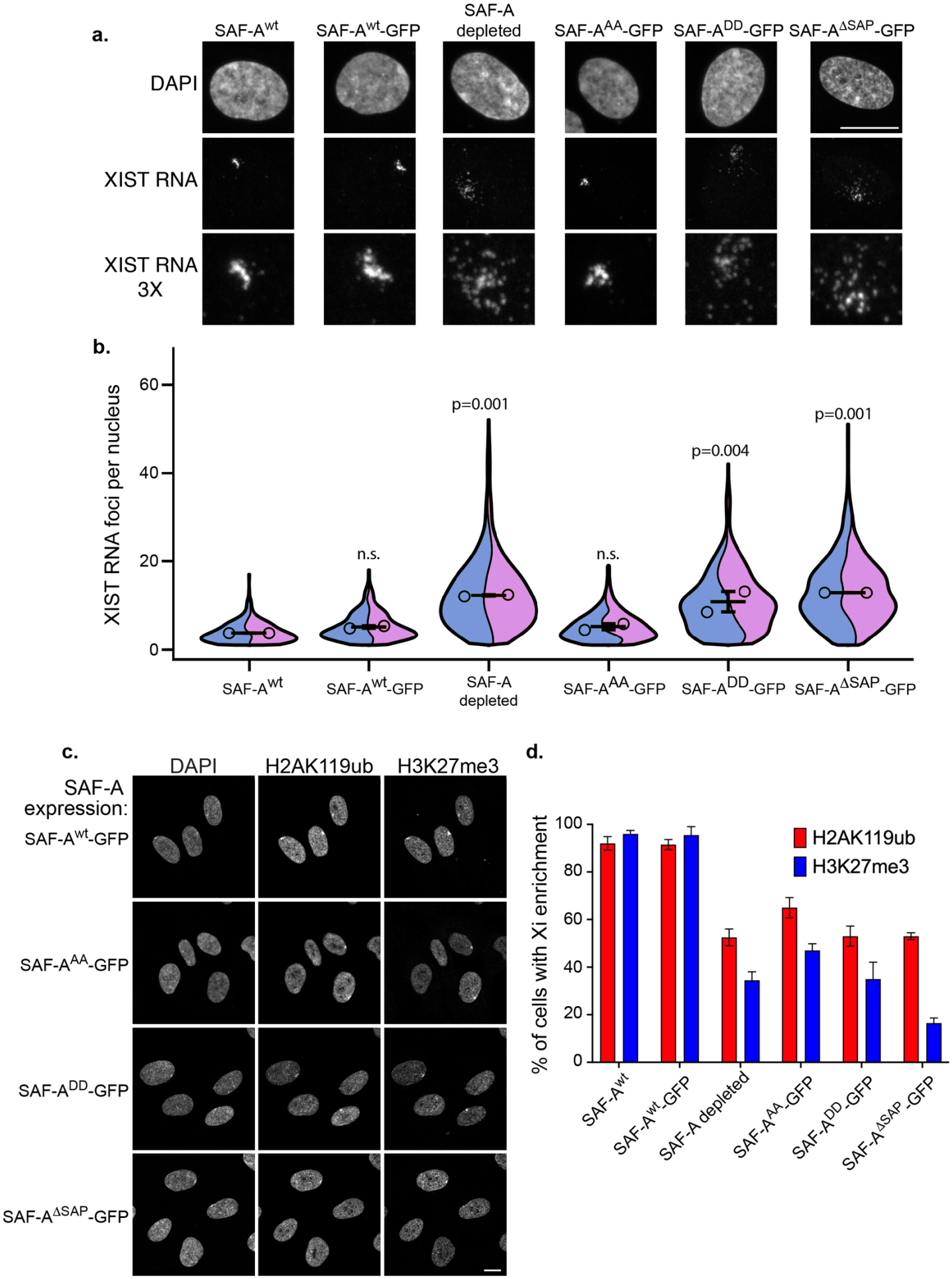
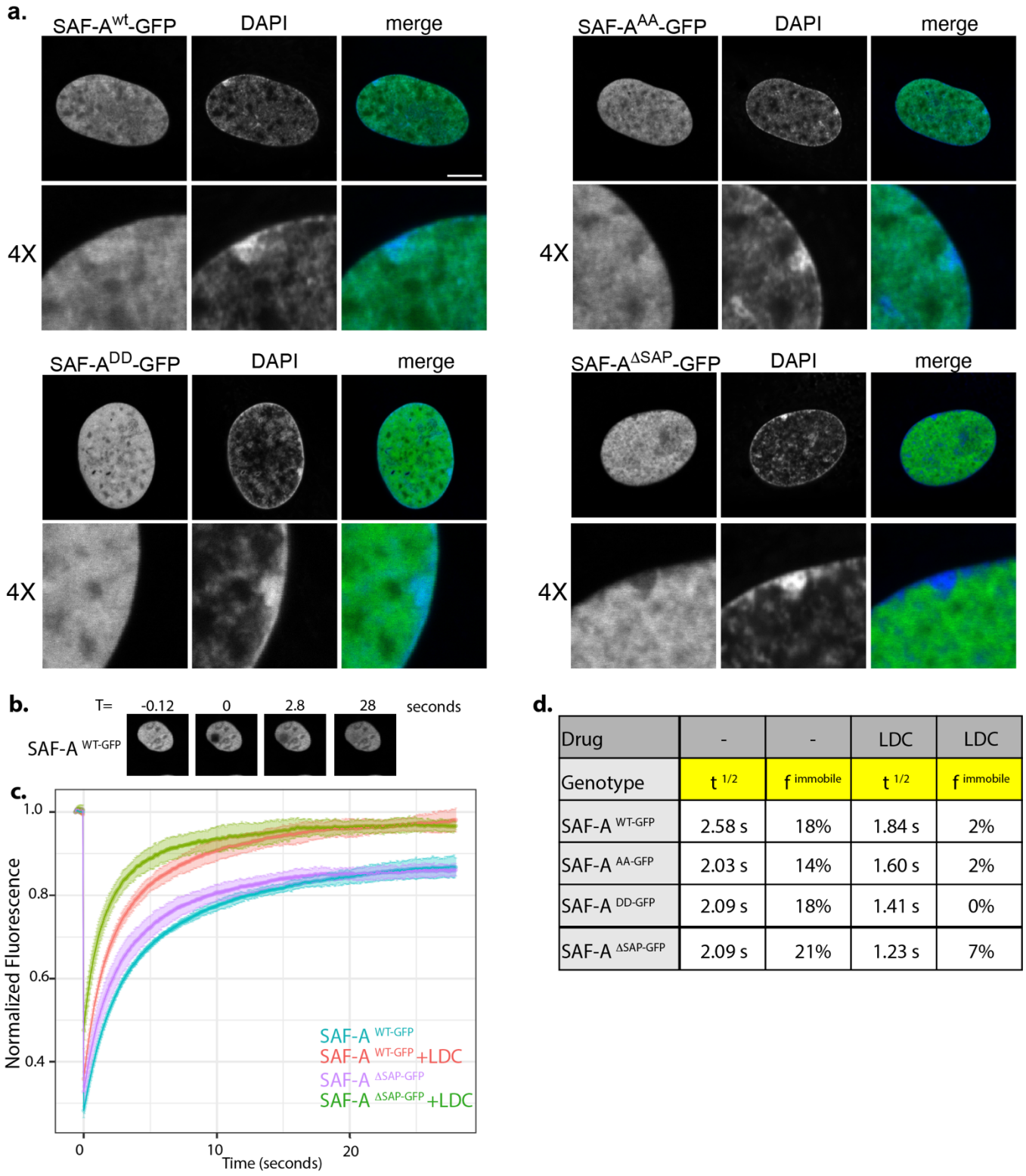

a.
SAF-A^wt
SAF-A^wt-GFP
SAF-A depleted
SAF-A^AA-GFP
SAF-A^DD-GFP
SAF-A^ΔSAP-GFP
DAPI
XIST RNA
XIST RNA 3X
b.
XIST RNA foci per nucleus
0
20
40
60
n.s.
n.s.
p=0.001
p=0.004
p=0.001
SAF-A^wt
SAF-A^wt-GFP
SAF-A depleted
SAF-A^AA-GFP
SAF-A^DD-GFP
SAF-A^ΔSAP-GFP
c.
DAPI
H2AK119ub
H3K27me3
SAF-A expression:
SAF-A^wt-GFP
SAF-A^AA-GFP
SAF-A^DD-GFP
SAF-A^ΔSAP-GFP
d.
% of cells with Xi enrichment
0
20
40
60
80
100
H2AK119ub
H3K27me3
SAF-A^wt
SAF-A^wt-GFP
SAF-A depleted
SAF-A^AA-GFP
SAF-A^DD-GFP
SAF-A^ΔSAP-GFP

Figure 2. The SAP domain and SAP domain serines S14 and 26 impact XIST RNA localization and PRC1/PRC2-dependent histone modifications on the Xi. A. XIST RNA FISH and DAPI staining in RPE-1 cells, and cells expressing SAF-A-GFP transgenes, 24 hours after treatment with doxycycline and IAA. Images are rendered as a maximum projection of a 3D stack. Bar, 10 μm. B. Quantitative measurement of XIST RNA foci per cell. Image stacks were acquired for at least 100 nuclei per genotype for two biological replicates and analyzed in Fiji software to count XIST RNA foci. Measurements are depicted as violin superplots. The average of each replicate is depicted as an open circle, whereas the average of both replicates is depicted as a horizontal line. The standard deviation of the two averages is shown. C. Immunofluorescence of histone H3K27me3 and H2AK119ub and DAPI staining, 24 hours after treatment with doxycycline and IAA. SAF-A-GFP allele expression is indicated to the left of the panel. Images are rendered as a maximum projection of a 3D stack. Bar, 10 μm. D. Quantitation of cell populations with H3K27me3 and H2AK119ub enrichment on the Xi. 100 cells were scored for H3K27me3 and H2AK119ub enrichment on the Xi in two biological replicates; the average and standard deviation is shown. Cell n for quantitation of XIST RNA particles is: RPE1 193, RPE2 211, SAF-A$^{WT\text{-}GFP1}$ 137, SAF-A$^{WT\text{-}GFP2}$ 236, SAF-A depletion 1 122, SAF-A depletion 2 220, , SAF-A$^{AA\text{-}GFP1}$ 170, , SAF-A$^{AA\text{-}GFP2}$ 206, , SAF-A$^{DD\text{-}GFP1}$ 133, , SAF-A$^{DD\text{-}GFP2}$ 198, SAF-A$^{\Delta SAP\text{-}GFP1}$ 160, , SAF-A$^{\Delta SAP\text{-}GFP2}$ 158. Statistical comparison of number of XIST RNA particles was performed using a one-way ANOVA followed by Tukey's tests with a Bonferroni correction.



Drug	-	-	LDC	LDC
Genotype	$t^{1/2}$	$f^{immobile}$	$t^{1/2}$	$f^{immobile}$
SAF-A $^{WT-GFP}$	2.58 s	18%	1.84 s	2%
SAF-A $^{AA-GFP}$	2.03 s	14%	1.60 s	2%
SAF-A $^{DD-GFP}$	2.09 s	18%	1.41 s	0%
SAF-A $^{\Delta SAP-GFP}$	2.09 s	21%	1.23 s	7%

Figure 3. The SAF-A SAP domain is important for nuclear dynamics and Xi localization. A. Live cell analysis of SAF-A^{wt}-GFP and SAP domain mutations as indicated. Cells were analyzed 24 hours after doxycycline and IAA treatment. Each image represents a single 0.2 µm slice. Bar, 10 µm. B. Images from a typical FRAP experiment using SAF-A^{wt}-GFP. C. FRAP recovery curves for SAF-A^{wt}-GFP, SAF-A^{AA}-GFP, SAF-A^{DD}GFP, and SAF-A$^{\Delta SAP}$-GFP, with and without transcriptional inhibition (LDC). Standard deviation of recovery is indicated in light colored error bars. D. Table of the $t^{1/2}$ recovery time and the immobile fraction for all SAF-A-GFP proteins, with and without transcriptional inhibition.

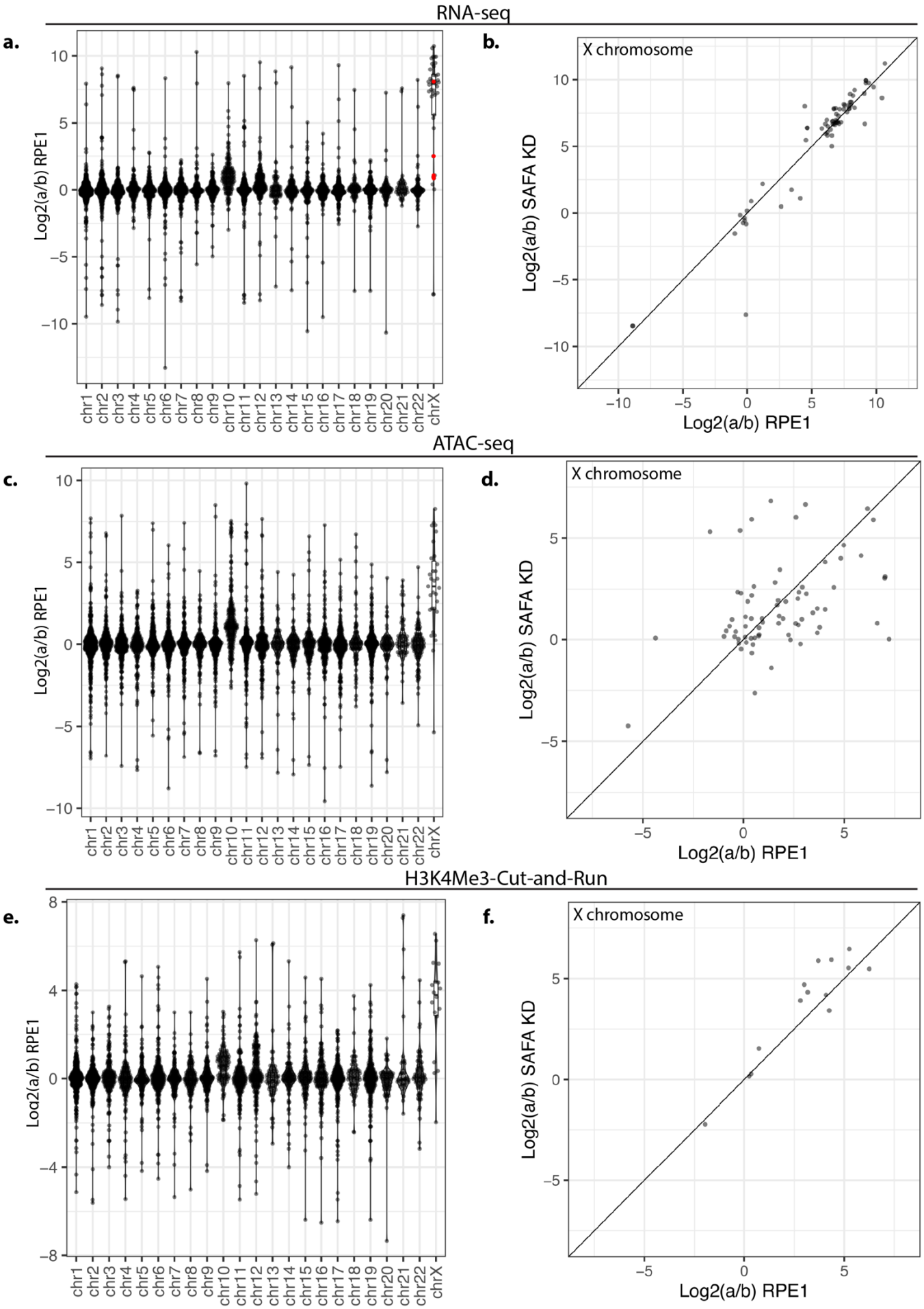
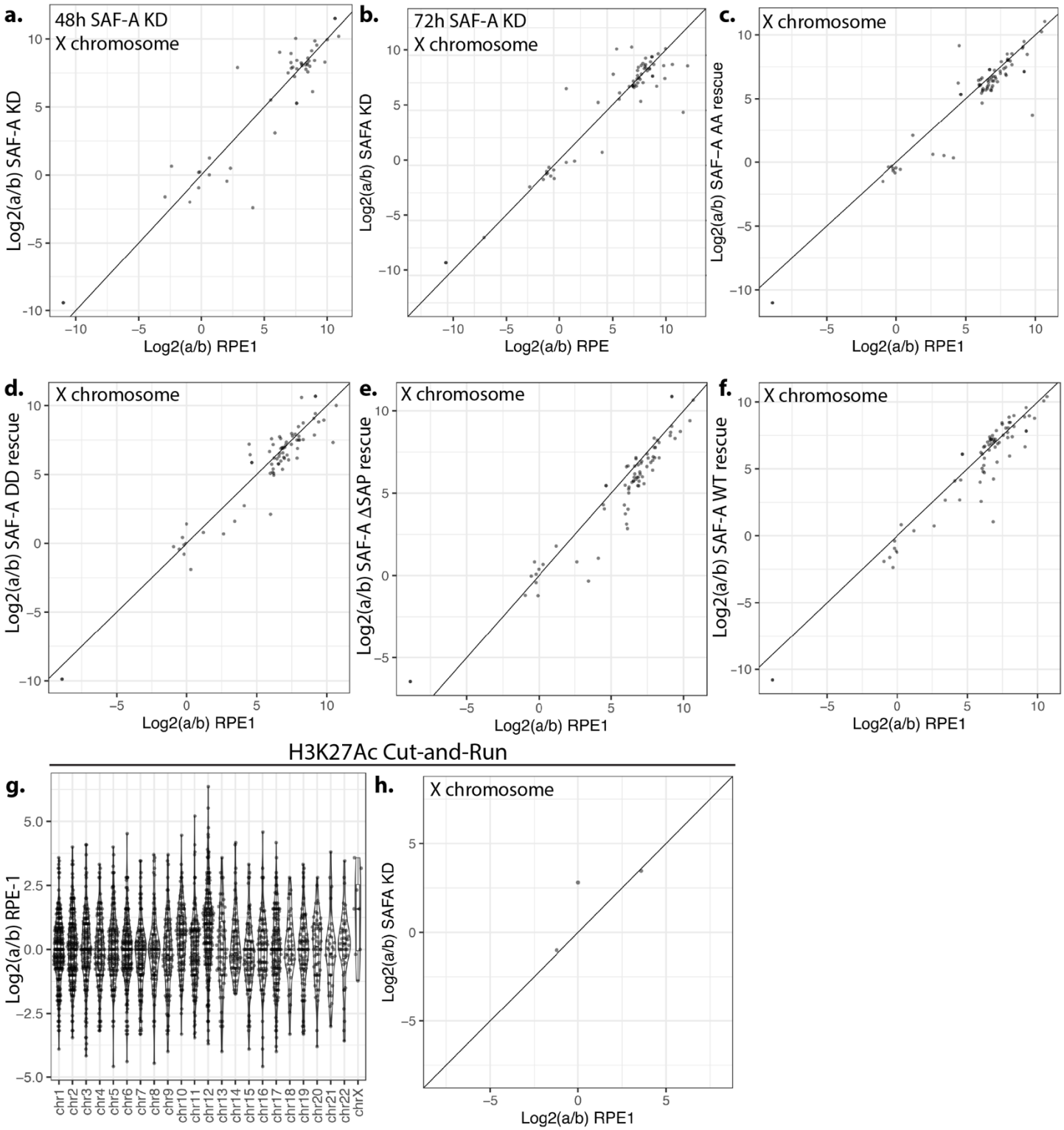

RNA-seq
a.
Log2(a/b) RPE1
b.
X chromosome
Log2(a/b) SAFA KD
Log2(a/b) RPE1
ATAC-seq
c.
Log2(a/b) RPE1
d.
X chromosome
Log2(a/b) SAFA KD
Log2(a/b) RPE1
H3K4Me3-Cut-and-Run
e.
Log2(a/b) RPE1
f.
X chromosome
Log2(a/b) SAFA KD
Log2(a/b) RPE1
chr1 chr2 chr3 chr4 chr5 chr6 chr7 chr8 chr9 chr10 chr11 chr12 chr13 chr14 chr15 chr16 chr17 chr18 chr19 chr20 chr21 chr22 chrX

Figure 4. SAF-A depletion does not reactivate gene expression on the inactive X chromosome. A. Allele-specific gene expression was calculated from RNA-seq libraries using a combination of PAC and edgeR in RPE-1 cells. ‘a to ‘b ratios are plotted by gene for each chromosome. B. Average a:b ratio for all genes on the X chromosome plotted for RPE-1 and SAF-A depleted cells. C. Allele-specific ATAC-seq was calculated using PAC and is plotted by gene for each chromosome. D. comparison of allele-specific ATAC-seq reads by gene plotted for RPE-1 cells and SAF-A depleted cells. E. Allele-specific Cut-and-Run for H3K4Me3 was calculated using PAC. ‘a/b’ ratio is plotted by gene for each chromosome. F. Allele-specific Cut-and-Run is plotted for RPE-1 cells and SAF-A depleted cells for X chromosome genes.



Supplemental Figure 1. Analysis of X-linked chromatin and gene expression in SAF-A mutants. A. Comparison of 'a/b' ratios for all X-linked genes after 48 hours of SAF-A KD. B. Comparison of 'a/b' ratios for all X-linked genes after 72 hours of SAF-A depletion. C.F. Comparison of 'a/b' ratios for all X-linked genes after in all SAF-A SAP domain mutants. G. Allele-specific Cut-and-Run was performed with H3K27Ac antibodies and calculated using PAC. 'a/b' ratios are plotted for each gene by chromosome. H. Comparison of 'a/b' ratios for X linked genes in SAF-A depleted cells.

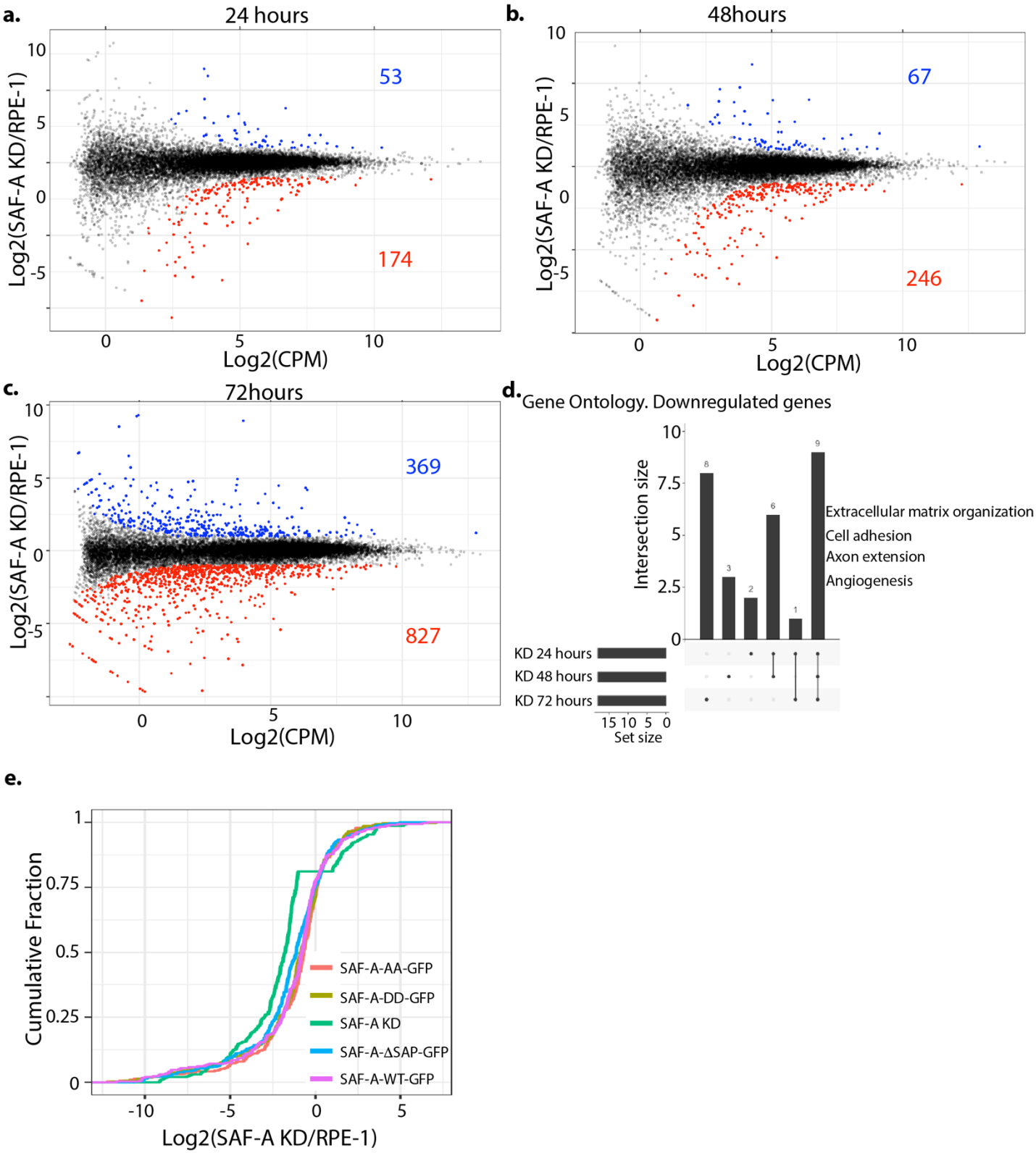
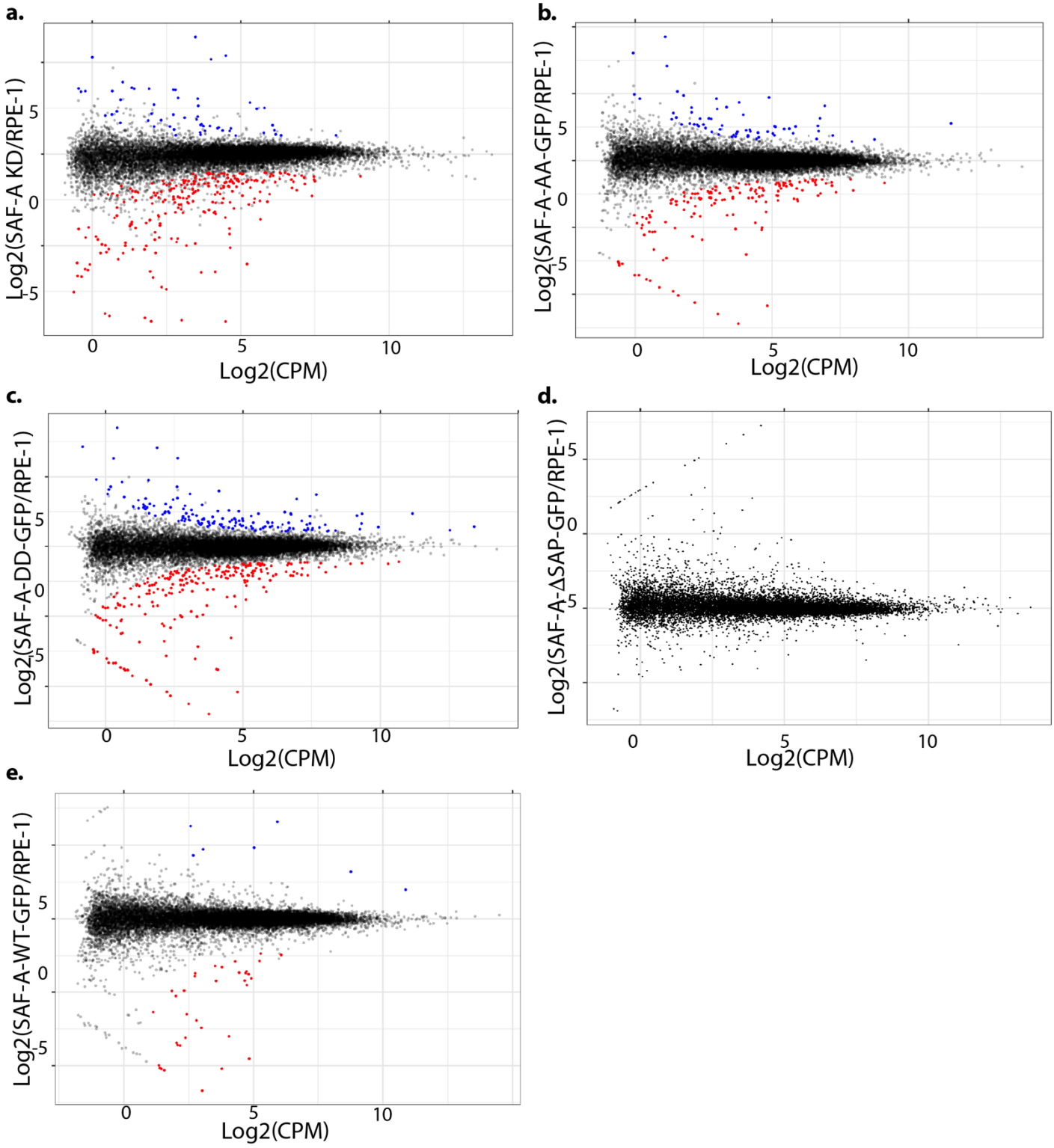


Figure 5. SAF-A depletion leads to minor defects in gene expression. A-C. SAF-A was depleted for 24, 48, or 72 hours through addition of auxin and gene expression was evaluated by RNA-seq and EdgeR. MD plots depict differentially expressed genes at each time point. Genes with an FDR < 0.01 are colored. d. Gene Ontology analysis of the genes downregulated following acute SAF-A depletion. e. Cumulative distribution plot depicting the magnitude of change in gene expression in each SAF-A domain mutant of the 227 genes significantly changed following SAF-A depletion (from A).



Supplemental Figure 2. SAF-A mutants do not alter gene expression at 24 hours. A-E. SAF-A was depleted by auxin addition and various mutants were expressed by the addition of doxycycline. Gene expression was evaluated at 24 hours using RNA-seq and EdgeR. MD plots depict significantly differentially expressed genes for each mutant (FDR < 0.01).

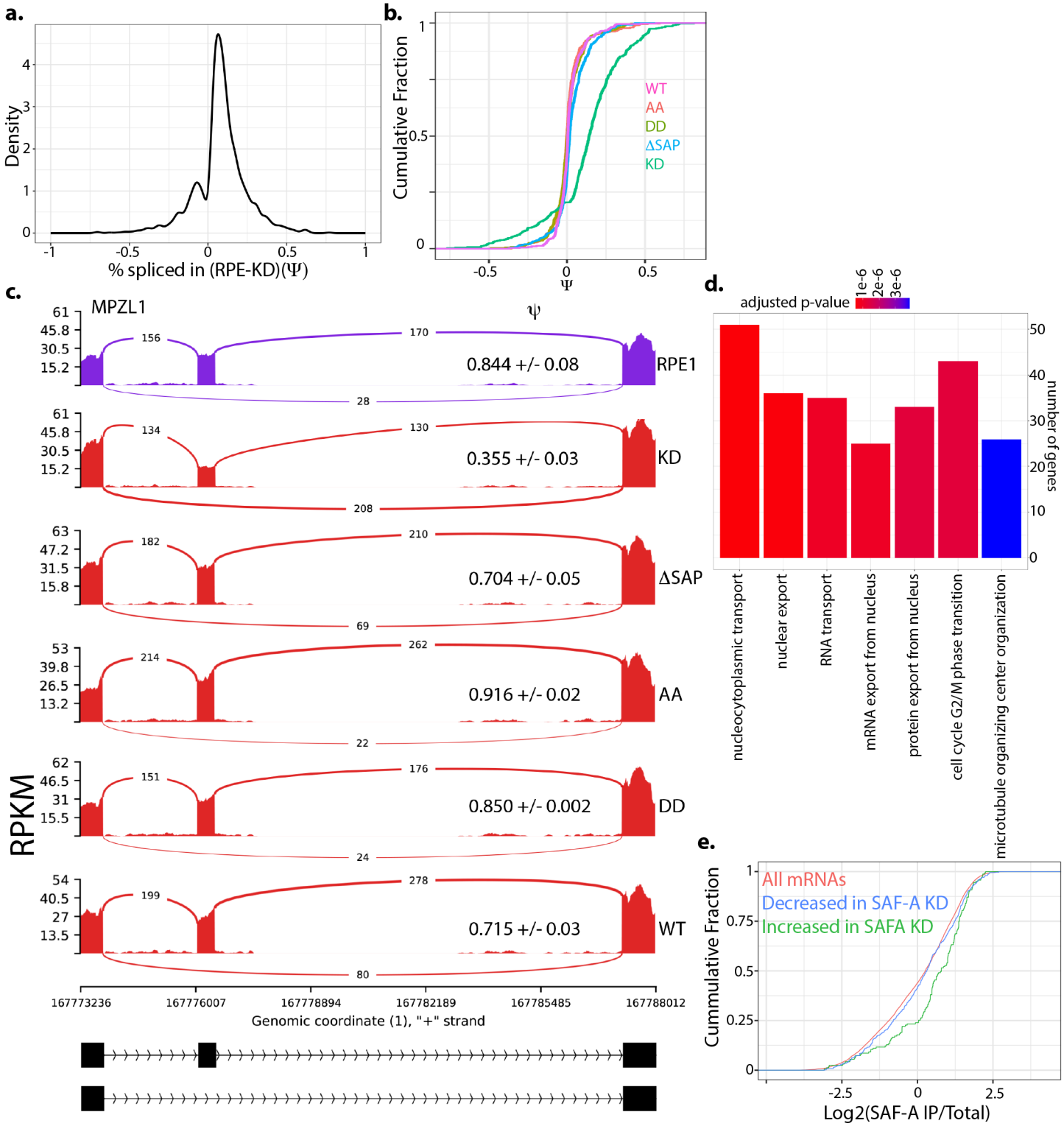
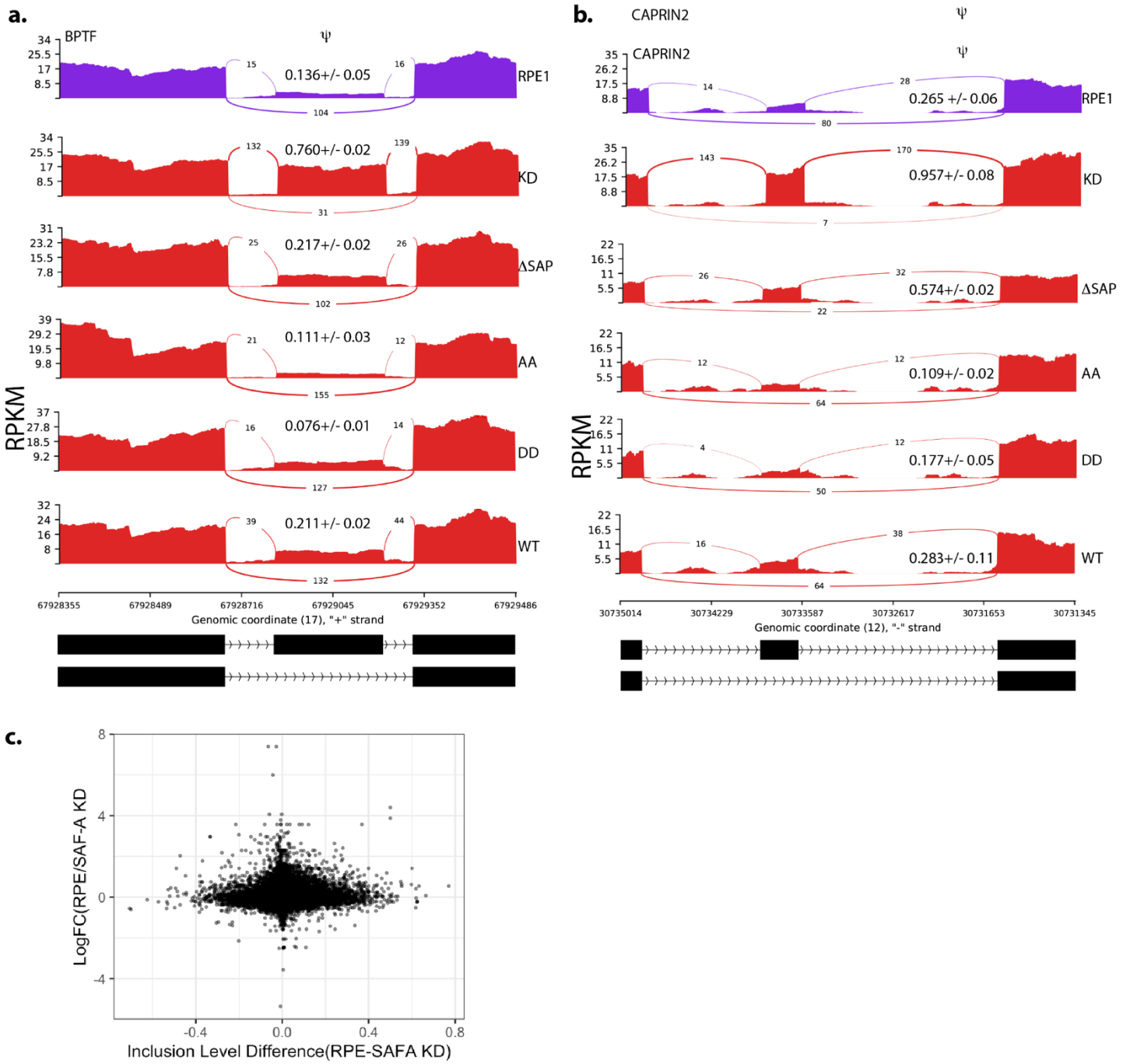
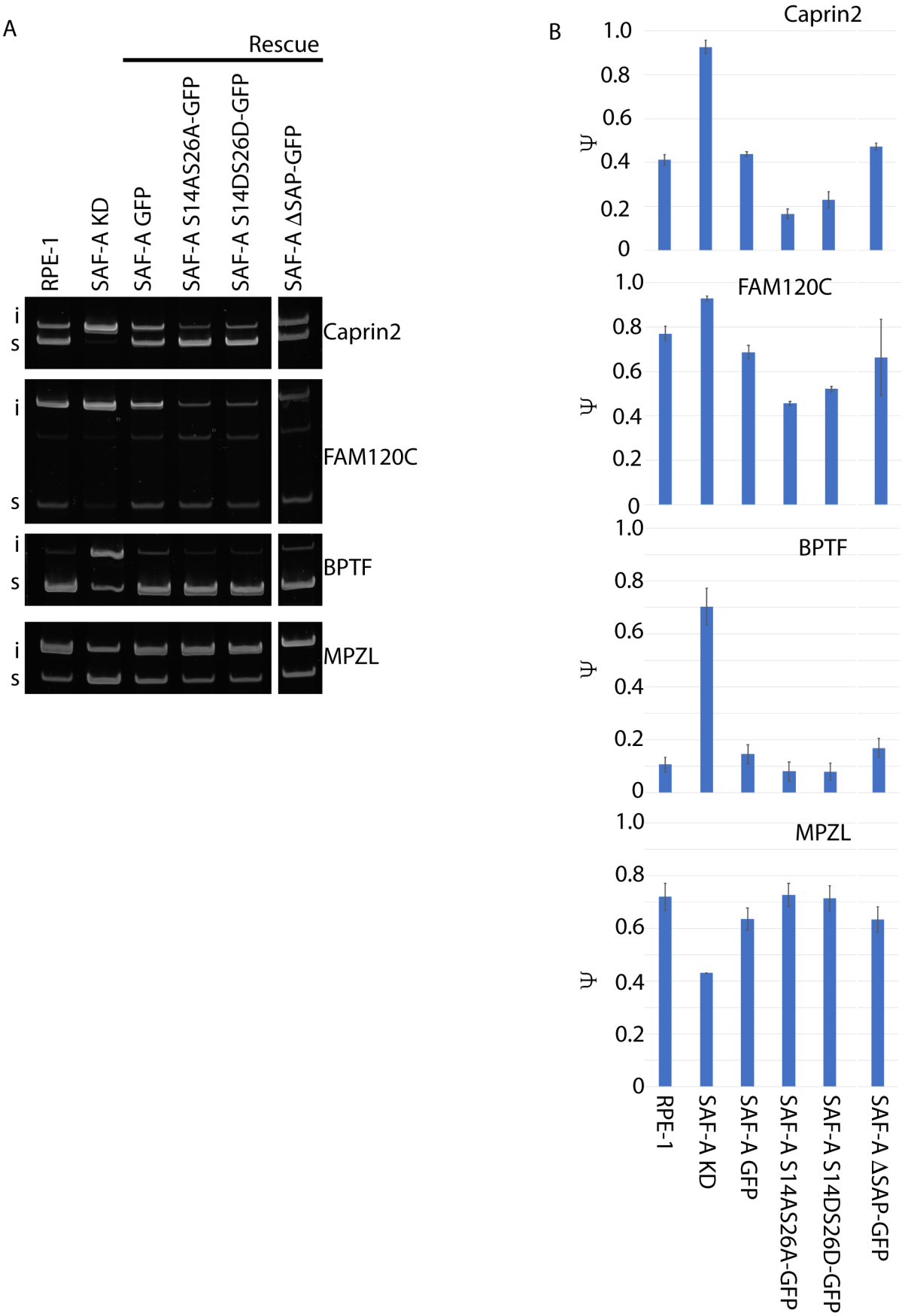


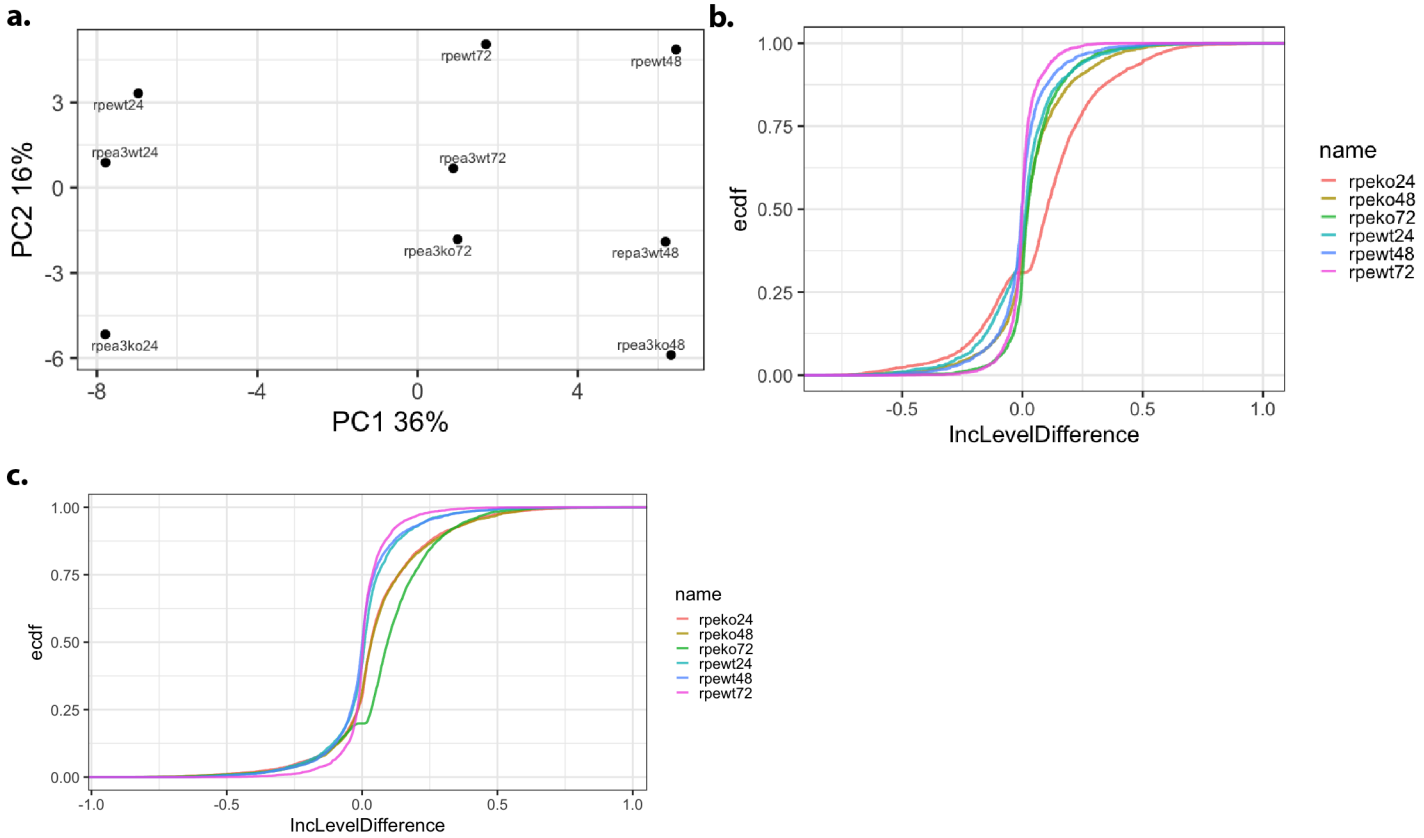
Figure 6. SAF-A depletion leads to widespread changes in mRNA splicing. A. mRNA splicing was evaluated in SAF-A depleted cells (24 hours) using rMATS. Density plot showing changes in exon inclusion. B. CDF plot illustrating the magnitude of changes in exon inclusion (of all exons significantly changed following SAF-A depletion) in all domain mutants. C. MISO plot showing an example altered exon. D. GO analysis of genes with altered splicing. E. CDF plot showing enrichment of all mRNAs, mRNAs with increased exon inclusion, and mRNAs with decreased exon inclusion in SAF-A RIP-seq libraries from interphase cells.



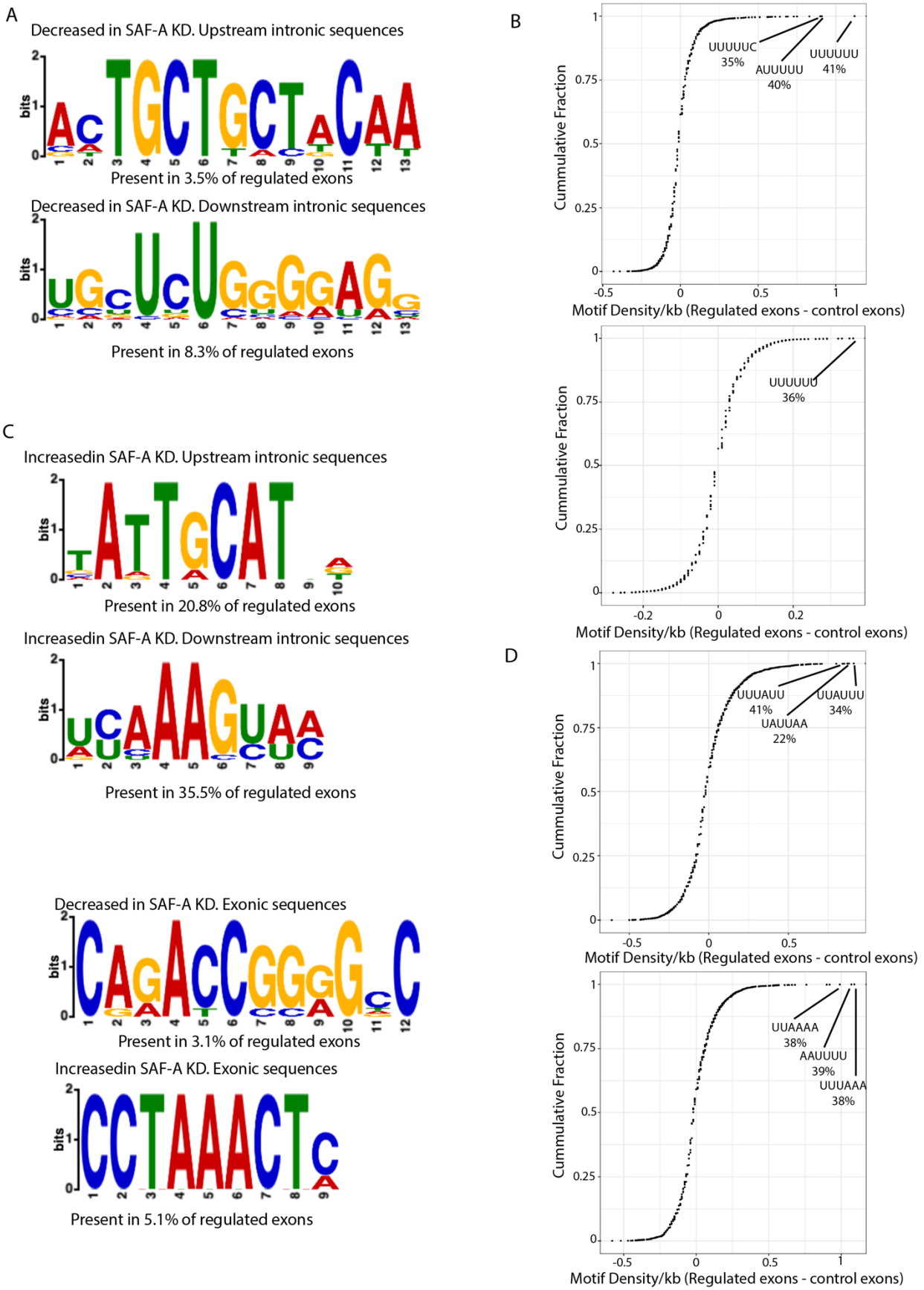
Supplemental Figure 3. MISO plots of additional regulated exons in each SAF-A mutant. A-B. MISO plots of SE events in WT RPE-1 cells, SAF-A depleted cells, and SAP domain mutants. C. Scatterplot comparison of changes in gene expression at 24h to changes in mRNA splicing.



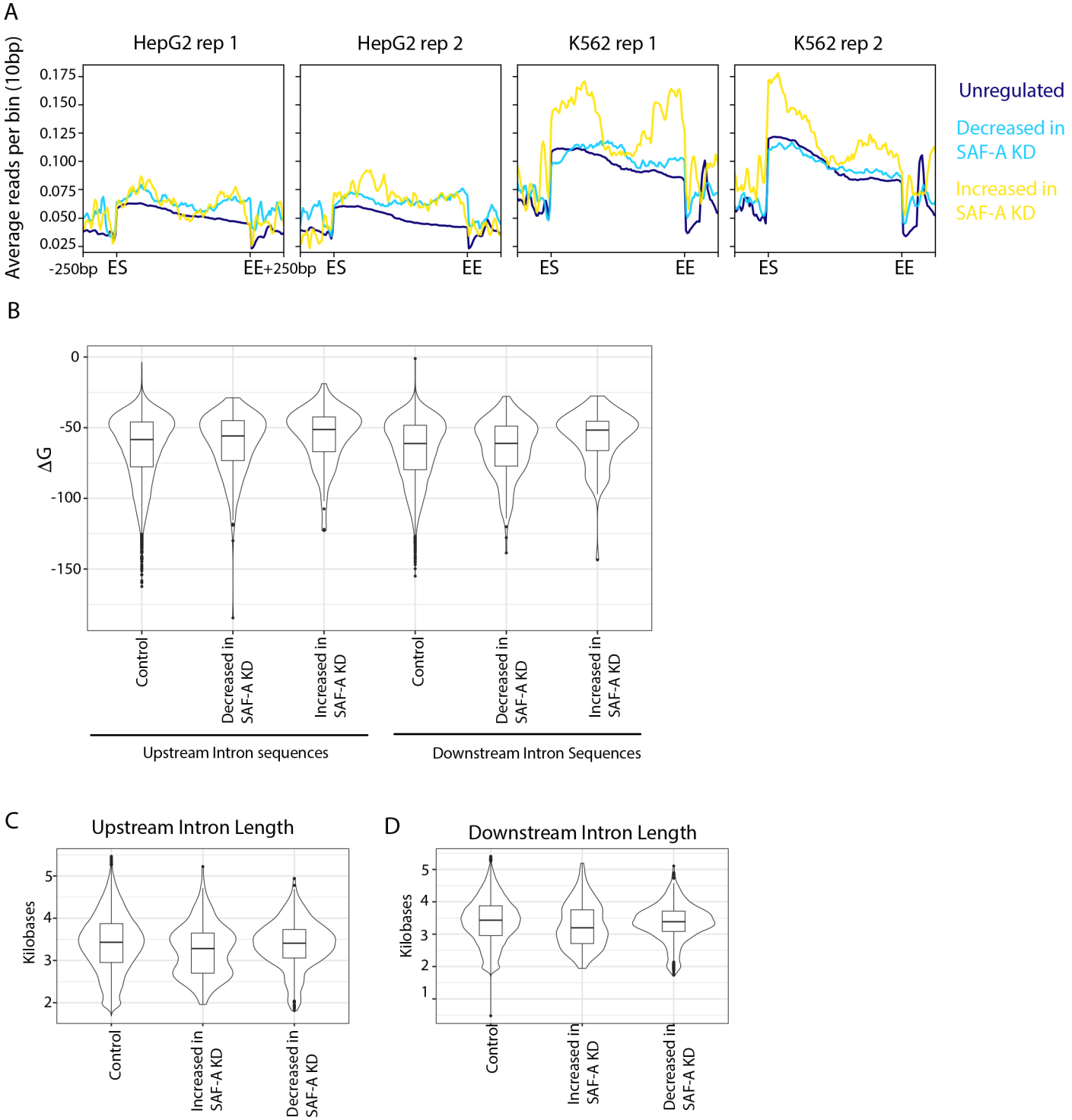
Supplemental Figure 4. Validation of predicted splicing changes using RT-PCR. A. PAGE gel analysis of exon inclusion for 4 different genes with predicted changes in exon inclusion in SAF-A depleted cells. B. Quantitation of percentage spliced in in each mutant from three biological replicates.



Supplemental Figure 5. Splicing defects do not increase with time of SAF-A depletion. A. PCA analysis of skipped exons in SAF-A KD, RPE-1, and SAF-A KD + SAF-A-GFP WT at each time point. B. CDF analysis of SE altered in SAF-A KD cells at 24 hours in each cell type at each time point. C. Same analysis as in B but with SE significantly altered in SAF-A KD at 72 hours.



Supplemental Figure 6. Motif analysis of SAF-A regulated exons. A. STREME analysis of upstream and downstream intronic sequences surrounding SAF-A regulated exons. B. Hexamer enrichment in the same sequences as in A. C. C-D STREME and hexamer analysis of sequences in and surrounding SAF-A regulated exons. Sequence sets tested are indicated on the plots.



Supplemental Figure 7. Analysis of SAF-A eCLIP data surrounding regulated exons. A. ENCODE project eCLIP data was analyzed using Deeptools in both HepG2 and K562 cells at SAF-A regulated exons (increased or decreased) and at unregulated genes. Average coverage and heatmap are depicted for all datasets. B. mFold calculation of ΔG for indicated sequenes. C-D. Calculated lengths of indicated sequence groups.

References

1. Skalska, L., et al., *Nascent RNA antagonizes the interaction of a set of regulatory proteins with chromatin.* Mol Cell, 2021. **81**(14): p. 2944-2959 e10.
2. Fackelmayer, F.O. and A. Richter, *Purification of two isoforms of hnRNP-U and characterization of their nucleic acid binding activity.* Biochemistry, 1994. **33**(34): p. 10416-22.
3. Romig, H., et al., *Characterization of SAF-A, a novel nuclear DNA binding protein from HeLa cells with high affinity for nuclear matrix/scaffold attachment DNA elements.* EMBO J, 1992. **11**(9): p. 3431-40. PMC:556878
4. Kiledjian, M. and G. Dreyfuss, *Primary structure and binding activity of the hnRNP U protein: binding RNA through RGG box.* EMBO J, 1992. **11**(7): p. 2655-64. PMC:556741
5. Renz, A. and F.O. Fackelmayer, *Purification and molecular cloning of the scaffold attachment factor B (SAF-B), a novel human nuclear protein that specifically binds to S/MAR-DNA.* Nucleic Acids Res, 1996. **24**(5): p. 843-9. PMC:PMC145707
6. Ye, J., et al., *hnRNP U protein is required for normal pre-mRNA splicing and postnatal heart development and function.* Proc Natl Acad Sci U S A, 2015. **112**(23): p. E3020-9. PMC:PMC4466706
7. Roshon, M.J. and H.E. Ruley, *Hypomorphic mutation in hnRNP U results in post-implantation lethality.* Transgenic Res, 2005. **14**(2): p. 179-92.
8. Tsherniak, A., et al., *Defining a Cancer Dependency Map.* Cell, 2017. **170**(3): p. 564-576 e16. PMC:PMC5667678
9. Dugger, S.A., et al., *Neurodevelopmental deficits and cell-type-specific transcriptomic perturbations in a mouse model of HNRNPU haploinsufficiency.* PLoS Genet, 2023. **19**(10): p. e1010952. PMC:PMC10569524 following competing interests. D.B.G. is a founder of Actio Biosciences, a founder of and holds equity in Praxis, serves as a consultant to AstraZeneca, and has received research support from Janssen, Gilead, Biogen, AstraZeneca and UCB. R.S.D serves as a consultant to AstraZeneca. S.C. serves a consultant for Q-State Biosciences, Inc.
10. Mastropasqua, F., et al., *Deficiency of the Heterogeneous Nuclear Ribonucleoprotein U locus leads to delayed hindbrain neurogenesis.* Biol Open, 2023. **12**(10). PMC:PMC10581386
11. Sapir, T., et al., *Heterogeneous nuclear ribonucleoprotein U (HNRNPU) safeguards the developing mouse cortex.* Nat Commun, 2022. **13**(1): p. 4209. PMC:PMC9304408 receives grant support from NINDS, NIMH, MURI, CDC, and NSF. He has equity and/or compensation from the following companies: Tilray, Receptor Life Sciences, Qstate Biosciences, Hitch Biosciences, Tevard Biosciences, Regel Biosciences, Script Biosciences, Actio Biosciences, Empatica, SilverSpike, and California Cannabis Enterprises (CCE). He has received consulting fees from Zogenix, Ultragenyx, BridgeBio, and Marinus. He holds patents for cannabidiol in treating neurological disorders, but these are owned by GW Pharmaceuticals, and he has waived any financial interests. He holds other patents in molecular biology. He is the managing partner of the PhiFund Ventures. Ingrid Scheffer has served on scientific advisory boards for BioMarin, Chiesi, Eisai, Encoded Therapeutics, GlaxoSmithKline, Knopp Biosciences, Nutricia, Rogcon, Takeda Pharmaceuticals, UCB, Xenon Pharmaceuticals; has received speaker honoraria from GlaxoSmithKline, UCB, BioMarin, Biocodex, Chiesi, Liva Nova and Eisai; has received funding for travel from UCB, Biocodex, GlaxoSmithKline, Biomarin and Eisai; has served as an investigator for Anavex Life Sciences, Cerecin Inc, Cereval Therapeutics, Eisai, Encoded Therapeutics, EpiMinder Inc, Epygenyx, ES-Therapeutics, GW Pharma, Marinus, Neurocrine BioSciences, Ovid Therapeutics, Takeda Pharmaceuticals, UCB, Ultragenyx, Xenon Pharmaceutical, Zogenix and Zynerba; and has consulted for Atheneum Partners, Care Beyond Diagnosis, Epilepsy Consortium, Ovid Therapeutics, UCB and Zynerba

Pharmaceuticals; and is a Non-Executive Director of Bellberry Ltd and a Director of the Australian Academy of Health and Medical Sciences and the Australian Council of Learned Academies Limited. The other authors declare no competing interests.

12. Huelga, S.C., et al., *Integrative genome-wide analysis reveals cooperative regulation of alternative splicing by hnRNP proteins.* Cell Rep, 2012. **1**(2): p. 167-78. PMC:3345519
13. Fan, H., et al., *The nuclear matrix protein HNRNPU maintains 3D genome architecture globally in mouse hepatocytes.* Genome Res, 2018. **28**(2): p. 192-202. PMC:PMC5793783
14. Nozawa, R.S., et al., *SAF-A Regulates Interphase Chromosome Structure through Oligomerization with Chromatin-Associated RNAs.* Cell, 2017. **169**(7): p. 1214-1227 e18. PMC:5473940
15. Hacisuleyman, E., et al., *Topological organization of multichromosomal regions by the long intergenic noncoding RNA Firre.* Nat Struct Mol Biol, 2014. **21**(2): p. 198-206. PMC:3950333
16. Hasegawa, Y., et al., *The matrix protein hnRNP U is required for chromosomal localization of Xist RNA.* Dev Cell, 2010. **19**(3): p. 469-76.
17. Sunwoo, H., et al., *Repeat E anchors Xist RNA to the inactive X chromosomal compartment through CDKN1A-interacting protein (CIZ1).* Proc Natl Acad Sci U S A, 2017. **114**(40): p. 10654-10659. PMC:PMC5635913
18. McHugh, C.A., et al., *The Xist lncRNA interacts directly with SHARP to silence transcription through HDAC3.* Nature, 2015. **521**(7551): p. 232-6.
19. Marenda, M., E. Lazarova, and N. Gilbert, *The role of SAF-A/hnRNP U in regulating chromatin structure.* Curr Opin Genet Dev, 2022. **72**: p. 38-44.
20. Kipp, M., et al., *SAF-Box, a conserved protein domain that specifically recognizes scaffold attachment region DNA.* Mol Cell Biol, 2000. **20**(20): p. 7480-9. PMC:PMC86301
21. Sharp, J.A., et al., *Cell division requires RNA eviction from condensing chromosomes.* J Cell Biol, 2020. **219**(11). PMC:PMC7549315
22. Miyaji, M., et al., *Selective DNA-binding of SP120 (rat ortholog of human hnRNP U) is mediated by arginine-glycine rich domain and modulated by RNA.* PLoS One, 2023. **18**(8): p. e0289599. PMC:PMC10403129
23. He, C., et al., *High-Resolution Mapping of RNA-Binding Regions in the Nuclear Proteome of Embryonic Stem Cells.* Mol Cell, 2016. **64**(2): p. 416-430. PMC:5222606
24. Hornbeck, P.V., et al., *PhosphoSitePlus, 2014: mutations, PTMs and recalibrations.* Nucleic Acids Res, 2015. **43**(Database issue): p. D512-20. PMC:PMC4383998
25. Chu, C., et al., *Systematic discovery of xist RNA binding proteins.* Cell, 2015. **161**(2): p. 404-16. PMC:4425988
26. Sakaguchi, T., et al., *Control of Chromosomal Localization of Xist by hnRNP U Family Molecules.* Dev Cell, 2016. **39**(1): p. 11-12.
27. Chadwick, B.P. and H.F. Willard, *Cell cycle-dependent localization of macroH2A in chromatin of the inactive X chromosome.* J Cell Biol, 2002. **157**(7): p. 1113-23. PMC:PMC2173542
28. Sunwoo, H., J.Y. Wu, and J.T. Lee, *The Xist RNA-PRC2 complex at 20-nm resolution reveals a low Xist stoichiometry and suggests a hit-and-run mechanism in mouse cells.* Proc Natl Acad Sci U S A, 2015. **112**(31): p. E4216-25. PMC:PMC4534268
29. Markaki, Y., et al., *Xist nucleates local protein gradients to propagate silencing across the X chromosome.* Cell, 2021. **184**(25): p. 6174-6192 e32. PMC:PMC8671326
30. Smeets, D., et al., *Three-dimensional super-resolution microscopy of the inactive X chromosome territory reveals a collapse of its active nuclear compartment harboring distinct Xist RNA foci.* Epigenetics Chromatin, 2014. **7**: p. 8. PMC:PMC4108088
31. Markaki, Y., et al., *The potential of 3D-FISH and super-resolution structured illumination microscopy for studies of 3D nuclear architecture: 3D structured illumination microscopy of defined chromosomal structures visualized by 3D (immuno)-FISH opens new perspectives for studies of nuclear architecture.* Bioessays, 2012. **34**(5): p. 412-26.

32. Kenny, M. and I. Schoen, *Violin SuperPlots: visualizing replicate heterogeneity in large data sets.* Mol Biol Cell, 2021. **32**(15): p. 1333-1334. PMC:PMC8694042
33. Jacobson, E.C., A. Pandya-Jones, and K. Plath, *A lifelong duty: how Xist maintains the inactive X chromosome.* Curr Opin Genet Dev, 2022. **75**: p. 101927. PMC:PMC9472561
34. Helbig, R. and F.O. Fackelmayer, *Scaffold attachment factor A (SAF-A) is concentrated in inactive X chromosome territories through its RGG domain.* Chromosoma, 2003. **112**(4): p. 173-82.
35. Pullirsch, D., et al., *The Trithorax group protein Ash2l and Saf-A are recruited to the inactive X chromosome at the onset of stable X inactivation.* Development, 2010. **137**(6): p. 935-43. PMC:PMC2834461
36. Creamer, K.M., H.J. Kolpa, and J.B. Lawrence, *Nascent RNA scaffolds contribute to chromosome territory architecture and counter chromatin compaction.* Mol Cell, 2021. **81**(17): p. 3509-3525 e5. PMC:PMC8419111
37. Albert, T.K., et al., *Characterization of molecular and cellular functions of the cyclin-dependent kinase CDK9 using a novel specific inhibitor.* Br J Pharmacol, 2014. **171**(1): p. 55-68. PMC:PMC3874696
38. Marenda, M., et al., *Nuclear RNA forms an interconnected network of transcription-dependent and tunable microgels.* bioRxiv, 2024: p. 2024.06.16.599208.
39. Brown, C.J. and H.F. Willard, *The human X-inactivation centre is not required for maintenance of X-chromosome inactivation.* Nature, 1994. **368**(6467): p. 154-6.
40. Csankovszki, G., A. Nagy, and R. Jaenisch, *Synergism of Xist RNA, DNA methylation, and histone hypoacetylation in maintaining X chromosome inactivation.* J Cell Biol, 2001. **153**(4): p. 773-84. PMC:PMC2192370
41. Loda, A., S. Collombet, and E. Heard, *Gene regulation in time and space during X-chromosome inactivation.* Nat Rev Mol Cell Biol, 2022. **23**(4): p. 231-249.
42. Tourdot, R.W., et al., *Determination of complete chromosomal haplotypes by bulk DNA sequencing.* Genome Biol, 2021. **22**(1): p. 139. PMC:PMC8101039
43. Saukkonen, A., H. Kilpinen, and A. Hodgkinson, *Highly accurate quantification of allelic gene expression for population and disease genetics.* Genome Res, 2022. **32**(8): p. 1565-72. PMC:PMC9435737
44. Volpe, E., et al., *The complete diploid reference genome of RPE-1 identifies human phased epigenetic landscapes.* bioRxiv, 2023. PMC:PMC10760208
45. Buenrostro, J.D., et al., *ATAC-seq: A Method for Assaying Chromatin Accessibility Genome-Wide.* Curr Protoc Mol Biol, 2015. **109**: p. 21 29 1-21 29 9. PMC:PMC4374986
46. Skene, P.J. and S. Henikoff, *An efficient targeted nuclease strategy for high-resolution mapping of DNA binding sites.* Elife, 2017. **6**. PMC:PMC5310842
47. Xiao, R., et al., *Nuclear matrix factor hnRNP U/SAF-A exerts a global control of alternative splicing by regulating U2 snRNP maturation.* Mol Cell, 2012. **45**(5): p. 656-68. PMC:PMC3299905
48. Lee, Y. and D.C. Rio, *Mechanisms and Regulation of Alternative Pre-mRNA Splicing.* Annu Rev Biochem, 2015. **84**: p. 291-323. PMC:PMC4526142
49. Bailey, T.L., et al., *The MEME Suite.* Nucleic Acids Res, 2015. **43**(W1): p. W39-49. PMC:PMC4489269
50. Yang, F., et al., *Genome-wide analysis identifies cis-acting elements regulating mRNA polyadenylation and translation during vertebrate oocyte maturation.* RNA, 2020. **26**(3): p. 324-344. PMC:PMC7025505
51. Guo, J.K., et al., *Denaturing purifications demonstrate that PRC2 and other widely reported chromatin proteins do not appear to bind directly to RNA in vivo.* Mol Cell, 2024. **84**(7): p. 1271-1289 e12. PMC:PMC10997485
52. Zuker, M., *Mfold web server for nucleic acid folding and hybridization prediction.* Nucleic Acids Res, 2003. **31**(13): p. 3406-15. PMC:PMC169194
53. Cho, N.H., et al., *OpenCell: Endogenous tagging for the cartography of human cellular organization.* Science, 2022. **375**(6585): p. eabi6983. PMC:PMC9119736

54. Fey, E.G., G. Krochmalnic, and S. Penman, *The nonchromatin substructures of the nucleus: the ribonucleoprotein (RNP)-containing and RNP-depleted matrices analyzed by sequential fractionation and resinless section electron microscopy.* J Cell Biol, 1986. **102**(5): p. 1654-65. PMC:PMC2114208
55. Fackelmayer, F.O., *A stable proteinaceous structure in the territory of inactive X chromosomes.* J Biol Chem, 2005. **280**(3): p. 1720-3.
56. Wutz, A., T.P. Rasmussen, and R. Jaenisch, *Chromosomal silencing and localization are mediated by different domains of Xist RNA.* Nat Genet, 2002. **30**(2): p. 167-74.
57. Sharp, J.A., et al., *Functional analysis of the microtubule-interacting transcriptome.* Mol Biol Cell, 2011. **22**(22): p. 4312-23. PMC:3216657
58. Xiao, C., et al., *The XIST noncoding RNA functions independently of BRCA1 in X inactivation.* Cell, 2007. **128**(5): p. 977-89.
59. Bancaud, A., et al., *Fluorescence perturbation techniques to study mobility and molecular dynamics of proteins in live cells: FRAP, photoactivation, photoconversion, and FLIP.* Cold Spring Harb Protoc, 2010. **2010**(12): p. pdb top90.
60. Chen, S., *Ultrafast one-pass FASTQ data preprocessing, quality control, and deduplication using fastp.* Imeta, 2023. **2**(2): p. e107. PMC:PMC10989850
61. Liao, Y., G.K. Smyth, and W. Shi, *The R package Rsubread is easier, faster, cheaper and better for alignment and quantification of RNA sequencing reads.* Nucleic Acids Res, 2019. **47**(8): p. e47. PMC:PMC6486549
62. Robinson, M.D., D.J. McCarthy, and G.K. Smyth, *edgeR: a Bioconductor package for differential expression analysis of digital gene expression data.* Bioinformatics, 2010. **26**(1): p. 139-40. PMC:PMC2796818
63. Chen, E.Y., et al., *Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool.* BMC Bioinformatics, 2013. **14**: p. 128. PMC:PMC3637064
64. Yu, G., et al., *clusterProfiler: an R package for comparing biological themes among gene clusters.* OMICS, 2012. **16**(5): p. 284-7. PMC:PMC3339379
65. Shen, S., et al., *rMATS: robust and flexible detection of differential alternative splicing from replicate RNA-Seq data.* Proc Natl Acad Sci U S A, 2014. **111**(51): p. E5593-601. PMC:PMC4280593
66. Ou, J., et al., *ATACseqQC: a Bioconductor package for post-alignment quality assessment of ATAC-seq data.* BMC Genomics, 2018. **19**(1): p. 169. PMC:PMC5831847