

## RESOURCE/METHODOLOGY

# An improved auxin-inducible degron system preserves native protein levels and enables rapid and specific protein depletion

Kizhakke Mattada Sathyan,<sup>1</sup> Brian D. McKenna,<sup>1</sup> Warren D. Anderson,<sup>2</sup> Fabiana M. Duarte,<sup>3</sup> Leighton Core,<sup>4</sup> and Michael J. Guertin<sup>1,2,5</sup>

<sup>1</sup>Biochemistry and Molecular Genetics Department, University of Virginia, Charlottesville, Virginia 22908, USA; <sup>2</sup>Center for Public Health Genomics, University of Virginia, Charlottesville, Virginia 22908, USA; <sup>3</sup>Department of Stem Cell and Regenerative Biology, Harvard University, Cambridge, Massachusetts 02138, USA; <sup>4</sup>Department of Molecular and Cell Biology, University of Connecticut, Storrs, Connecticut 06269, USA; <sup>5</sup>Cancer Center, University of Virginia, Charlottesville, Virginia 22908, USA

**Rapid perturbation of protein function permits the ability to define primary molecular responses while avoiding downstream cumulative effects of protein dysregulation. The auxin-inducible degron (AID) system was developed as a tool to achieve rapid and inducible protein degradation in nonplant systems. However, tagging proteins at their endogenous loci results in chronic auxin-independent degradation by the proteasome. To correct this deficiency, we expressed the auxin response transcription factor (ARF) in an improved inducible degron system. ARF is absent from previously engineered AID systems but is a critical component of native auxin signaling. In plants, ARF directly interacts with AID in the absence of auxin, and we found that expression of the ARF PB1 (Phox and Bem1) domain suppresses constitutive degradation of AID-tagged proteins. Moreover, the rate of auxin-induced AID degradation is substantially faster in the ARF-AID system. To test the ARF-AID system in a quantitative and sensitive manner, we measured genome-wide changes in nascent transcription after rapidly depleting the ZNF143 transcription factor. Transcriptional profiling indicates that ZNF143 activates transcription in *cis* and regulates promoter-proximal paused RNA polymerase density. Rapidly inducible degradation systems that preserve the target protein's native expression levels and patterns will revolutionize the study of biological systems by enabling specific and temporally defined protein dysregulation.**

[**Keywords:** auxin-inducible degron; auxin response factor; ZNF143; transcription factors; RNA polymerase pausing]

Supplemental material is available for this article.

Received May 21, 2019; revised version accepted July 18, 2019.

The function of proteins can be studied in cells using RNAi depletion, loss-of-function mutants, temperature-sensitive mutations, small molecule inhibitors, CRISPR interference, or nucleic acid aptamers. The two greatest limitations of these methods are (1) the chronic nature of permanently disrupting function and (2) the limited availability and specificity of inhibitor molecules. Inducible degradation methods, such as auxin-inducible degron (AID) (Nishimura et al. 2009) or degradation tag (dTAG) (Nabet et al. 2018) systems, can overcome these limitations. Rapidly regulated systems and inhibitors permit measurements of primary changes in molecular, cellular, and organismal phenotypes and subsequent tracking of

the cascade that accompanies immediate protein dysregulation. Gene-editing techniques now permit endogenous tagging of genes, which should preserve the target gene's expression levels and patterns. Endogenous fusion of target genes with rapidly inducible degradation tags have the potential to revolutionize the way we study biological systems by providing specific and temporally defined perturbation techniques for any protein of interest.

Auxin is a plant hormone that regulates various aspects of plant growth and development. In plants, auxin signaling triggers a rapid switch between transcriptional repression and transcriptional activation (Supplemental Fig. S1; Calderon-Villalobos et al. 2010; Lavy and Estelle 2016; Li et al. 2016). There are three key components in this signal

**Correspondence:** [guertin@virginia.edu](mailto:guertin@virginia.edu)

Article published online ahead of print. Article and publication date are online at <http://www.genesdev.org/cgi/doi/10.1101/gad.328237.119>. Freely available online through the *Genes & Development* Open Access option.

© 2019 Sathyan et al. This article, published in *Genes & Development*, is available under a Creative Commons License (Attribution 4.0 International), as described at <http://creativecommons.org/licenses/by/4.0/>.

transduction system: (1) transport inhibitor response 1 (TIR1), a subunit of a ubiquitin ligase complex that binds to the target substrate; (2) the auxin response transcription factors (ARF), which directly regulate gene expression; and (3) auxin/indole-3-acetic acid (Aux/IAA) proteins, which are destabilized in the presence of auxin-mediated ubiquitination. In the absence of auxin, domains III and IV of the Aux/IAA family of proteins form heterodimeric complexes with the PB1 (Phox and Bem1) domain of ARF proteins (Kim et al. 1997; Ulmasov et al. 1999). Domain I of Aux/IAA interacts with a plant-specific transcriptional corepressor, Topless (TPL) (Szemenyei et al. 2008). The repressive function of TPL dominates relative to the activation function of ARF transcription factors (Tiwari et al. 2004). In the presence of auxin, TIR1 interacts with domain II of Aux/IAA to facilitate ubiquitination and degradation of Aux/IAA (Gray et al. 2001; Dharmasiri et al. 2003), which liberates ARFs to regulate transcription. This AID system was engineered to function outside plant cells (Nishimura et al. 2009).

Although the AID system has been widely adopted to degrade tagged proteins of interest, recent studies report auxin-independent depletion of endogenously tagged proteins (Morawska and Ulrich 2013; Nishimura and Fukagawa 2017; Zasadzińska et al. 2018). For instance, tagging the centromeric histone chaperone HJURP results in depletion of >90% of HJURP protein in the absence of auxin in human cell lines (Zasadzińska et al. 2018). Chronic depletion has also been reported in chicken cells (Nishimura and Fukagawa 2017) and yeast cells (Morawska and Ulrich 2013), suggesting that auxin-independent depletion may be a systemic problem when tagging endogenous genes. The extent of factor-dependent depletion is often impossible to evaluate because many studies do not directly compare protein levels in the progenitor and tagged cell lines (McKinley et al. 2015; Hoffmann et al. 2016; Cao et al. 2018). Neither defining a minimal degradation domain of AID nor using an inducible TIR1 system has overcome this deficiency (Morawska and Ulrich 2013; Samejima et al. 2014; Mendoza-Ochoa et al. 2019). In yeast, auxin-independent depletion of the AID-tagged factors is dependent on cellular TIR1 concentration, suggesting that depletion is due to ubiquitin-mediated degradation (Mendoza-Ochoa et al. 2019).

Here, we show that AID tagging of endogenous genes commonly results in auxin-independent and chronic depletion by proteasome-mediated protein degradation. We found that coexpressing the PB1 domain of ARF improved the robustness of the AID system by rescuing auxin-independent degradation and increasing the rate of auxin-induced degradation. Our control experiments also revealed that auxin treatment alone results in the activation of the aryl hydrocarbon receptor (AHR) transcription factor and regulation of AHR target genes. Excluding AHR-responsive genes from downstream differential expression analyses is critical when investigating the activity of AID-tagged transcriptional regulators. Collectively, these improvements enhance the robustness, sensitivity, and specificity of the AID system. We applied ARF-rescued AID-mediated rapid degra depletion to the tran-

scription factor ZNF143 to identify the primary ZNF143 response genes and define a functional role of ZNF143 in transcriptional regulation.

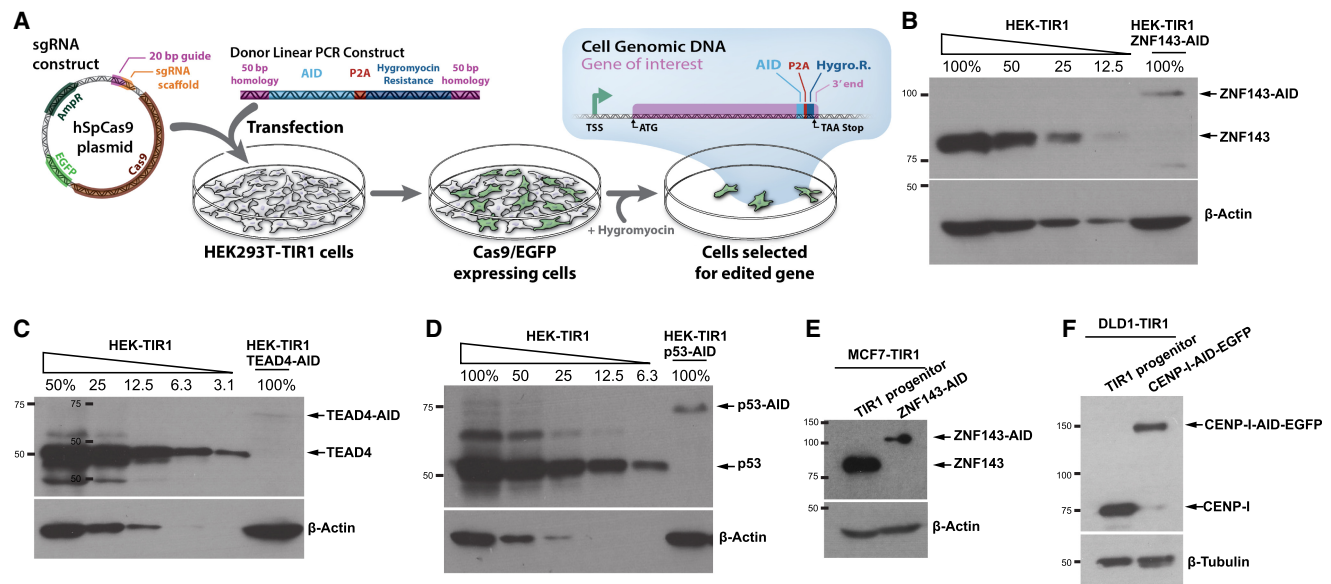
## Results

### *Endogenous AID tagging results in chronic target protein depletion*

In nonplant systems, the auxin-mediated degradation system requires the presence of exogenously expressed TIR1 (Nishimura et al. 2009). We generated an HEK293T cell line with the TIR1 gene stably integrated into the AAVS1 locus (Supplemental Fig. S2; Mali et al. 2013; Natsume et al. 2016). We independently tagged all copies of either ZNF143, TEAD4, or p53 in these progenitor HEK293T-TIR1 cells (Fig. 1A). In the absence of auxin, tagging the factors resulted in depletion of the proteins to levels that range from between <3% and 15% of endogenous levels, as measured by quantitative Western blots (Fig. 1B–D). To determine whether chronic depletion is unique to HEK293T cells, we generated an MCF-7 cell line with TIR1 incorporated heterozygously into the AAVS1 locus and tagged all copies of ZNF143, which also resulted in auxin-independent ZNF143 depletion (Fig. 1E). Next, we characterized a previously published AID-tagged CENP-I colorectal epithelial DLD-1 cell line (McKinley et al. 2015). CENP-I is depleted to ~50% of the abundance of the progenitor DLD-1-TIR1 cell line (Fig. 1F). The limited degree of CENP-I depletion relative to ZNF143, TEAD4, and p53 may be because CENP-I is an essential protein, and a minimal abundance of CENP-I is necessary for cell line viability. Basal depletion of these factors in three distinct cell lines confirms the generality of auxin-independent depletion of endogenously tagged proteins.

### *Proteasome-mediated degradation drives auxin-independent depletion of proteins*

We next sought to investigate the mechanism by which AID-tagged proteins are depleted by focusing on AID-tagged ZNF143. We performed nascent RNA sequencing (Core et al. 2008; Kwak et al. 2013; Mahat et al. 2016) in the progenitor HEK293T-TIR1 cells and the ZNF143-AID cells. Nascent ZNF143 transcript levels remain unchanged (1.08-fold increase; false discovery rate [FDR] = 0.27) in the AID-tagged ZNF143 cell line (Fig. 2A). Because the auxin-inducible system is dependent on the proteasome and because AID can interact with TIR1 in the absence of auxin (Gray et al. 2001; Dharmasiri et al. 2005; Kepinski and Leyser 2005; Tan et al. 2007), we hypothesized that low protein abundance may be mediated by an auxin-independent interaction with TIR1, ubiquitination, and subsequent proteasome-mediated degradation. We treated ZNF143-AID cells with the proteasome inhibitor MG132 and observed a modest increase in ZNF143 levels after 4.5 h (Supplemental Fig. S3A). TIR1 depletion also results in higher ZNF143-AID and TEAD4-AID levels (Supplemental Fig. S3B,C). These results indicate that chronic auxin-independent depletion



**Figure 1.** Endogenous tagging of genes with AID results in auxin-independent protein depletion. (A) Genes of interest were targeted by cotransfecting sgRNAs (Supplemental Table S2) targeting the 3' end of the coding sequence with a homology-directed repair donor construct. The donor construct contained the AID domain fused to a hygromycin resistance marker; an intervening porcine teschovirus-1 2A (P2A) site liberates the hygromycin resistance marker during protein translation. This construct was flanked by 50-base-pair (bp) homology regions that correspond to the sequences flanking the sgRNA recognition sequence and Cas9 cleavage site. (B–D) We quantified AID-tagged protein abundance with quantitative Western blots and a dilution of the progenitor cell line for the standard curve. Endogenous homozygous tagging of ZNF143 (B), TEAD4 (C), and p53 (D) in HEK293T cells results in auxin-independent chronic protein depletion. MCF-7 (E) and DLD-1 (F) cell lines also exhibit depletion of endogenously tagged ZNF143 (E) and CENP-I (F).

of AID-tagged proteins is TIR1-mediated and due to proteasomal degradation (Supplemental Fig. S3D).

#### ARF rescues auxin-independent proteasomal degradation

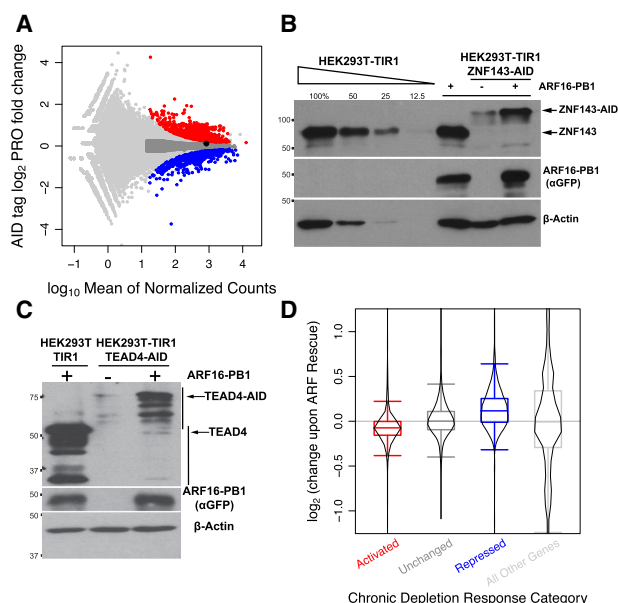
TIR1 and AID proteins can directly interact in the absence of auxin *in vitro* (Dharmasiri et al. 2005; Kepinski and Leyser 2005; Tan et al. 2007). However, degradation is tightly and robustly regulated in plants (Supplemental Fig. S1; Chapman and Estelle 2009). We proposed that expression of an ARF protein, which is an AID interaction partner, may confer protection of the tagged protein from auxin-independent ubiquitination and degradation. To test this hypothesis, we supplemented the engineered AID system by expressing the AID-interacting PB1 domain of ARF. We reconstituted the system by expressing the PB1 domain of *Oryza sativa* ARF16 and ARF25 based on yeast two-hybrid experiments that quantified interaction of these ARFs with IAA proteins (Shen et al. 2010). Moreover, ARF16 and ARF25 harbor conserved charged residues, corresponding to K944, D994, and D998 of ARF16 (Wang et al. 2007), at critical positions within the ARF/IAA-binding interface (Korasick et al. 2014; Nanao et al. 2014). Transfection of either eGFP-ARF25-MR-PB1 or eGFP-ARF16-PB1 stabilizes TEAD4-AID, with ARF16 promoting a higher degree of TEAD4 stability (Supplemental Fig. S4A,B). These results prompted us to generate HEK293T-TIR1-ZNF143-AID cells with stable genetic integration and expression of ARF16-PB1. This strategy restored ZNF143

levels to >50% of untagged levels (Fig. 2B; Supplemental Fig. S4C,D). Similarly, we found that genetic integration and expression of eGFP-ARF16-PB1 stabilized endogenously tagged TEAD4-AID (Fig. 2C). In contrast, stable expression of eGFP-ARF16-PB1 did not alter ZNF143, TEAD4, or p53 protein levels in HEK293T or HEK293T-TIR1 cells (Supplemental Fig. S4E).

Transcriptional output is a quantitative measure of ARF-mediated functional rescue of ZNF143. We performed nascent RNA transcriptional profiling (Core et al. 2008) with the three successive genetically modified HEK293T cells: progenitor TIR1 cells, chronic ZNF143-depleted AID-tagged cells, and ARF-rescued ZNF143-AID cells. Chronic ZNF143 depletion resulted in activation of 1188 genes and repression of 774 genes at a FDR of 0.001 (Fig. 2A). Next, we analyzed the raw changes in expression upon ARF rescue to determine whether rescuing ZNF143 stability can functionally rescue gene expression profiles. Of the 1188 genes activated upon chronic ZNF143 depletion, 899 (76%) decrease their expression upon ARF rescue (Fig. 2D). Of the 774 chronically repressed genes, 561 (72%) increase their expression upon ARF rescue (Fig. 2D). These changes are consistent with a functional rescue of gene expression upon ARF rescue of ZNF143 stability.

#### ARF rescue mediates rapid auxin-inducible degradation

We treated ZNF143-AID and TEAD4-AID cells with 500  $\mu$ M auxin to determine how the ARF16 rescue affects



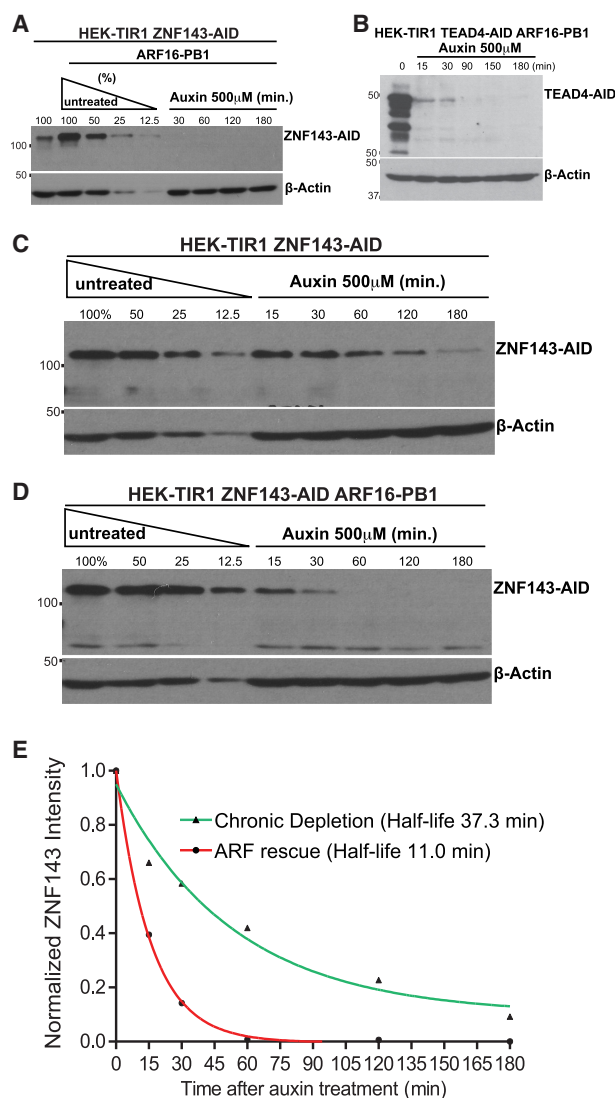
**Figure 2.** Expressing the PB1 domain of ARF rescues the proteasome-dependent chronic depletion of AID-tagged proteins. (A) Red genes are activated upon chronic ZNF143 depletion, blue genes are repressed, dark-gray genes are confidently unchanged and expressed at levels comparable with those of changed genes, and light-gray points represent all other genes. ZNF143 (black point) expression is unaffected. (B) Stable expression of the ARF-PB1 domain in HEK293T-TIR1 cells did not change the protein levels of ZNF143 (fifth lane) but increased the stability of ZNF143-AID (seventh lane). (C) TEAD4 levels also increase upon ARF16 expression. (D) Genes that are activated upon chronic ZNF143 depletion in A are, on average, repressed upon ARF rescue, and genes that are repressed upon chronic depletion are activated upon ARF rescue.

inducible depletion. Auxin treatment induces degradation of both ZNF143-AID and TEAD4-AID in a time-dependent manner (Fig. 3A,B). Importantly, the rate of degradation of ZNF143 was increased upon ARF16-PB1 rescue when compared with cells not rescued with ARF16-PB1 (Fig. 3C–E). To test whether ARF16-PB1 rescue affects the synthesis rate of ZNF143-AID and thus contributes to the perceived degradation rate, we simultaneously blocked new protein synthesis with cycloheximide at the time of auxin treatment (Supplemental Fig. S5). Upon blocking protein synthesis, the ZNF143-AID protein degraded faster in the presence of ARF16-PB1 (Supplemental Fig. S5). Therefore, ARF expression promotes faster degradation kinetics and does not influence protein synthesis rate.

#### ARF interacts with AID to rescue AID tag stability in mammalian cells

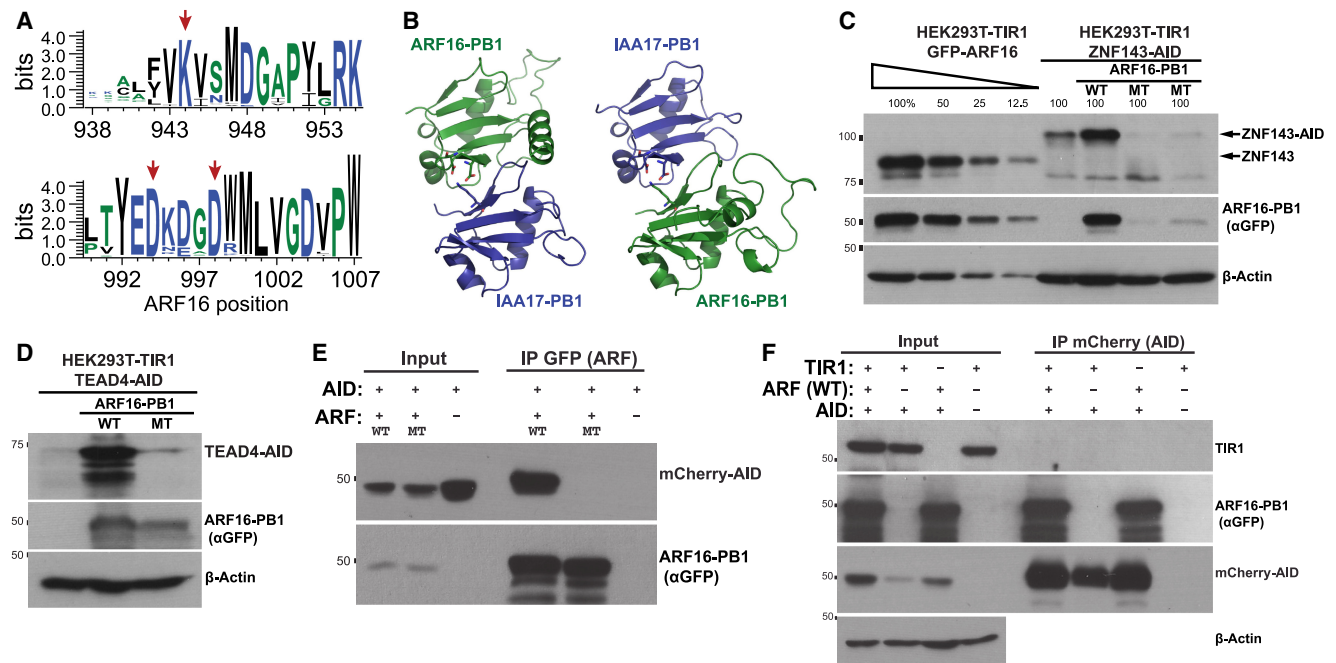
To test the specificity of the ARF16-PB1 rescue, we mutated ARF16 residues within the interaction domain interface that are critical for its interaction with IAA17 (Fig. 4A,B; Korasick et al. 2014; Nanao et al. 2014). We converted K944, D994, and D998 to alanine in the

ARF16-PB1-MT. In *Arabidopsis thaliana*, the corresponding mutations abolish ARF homodimerization and heterodimerization with IAA17 (Korasick et al. 2014). We tested whether this mutant is capable of rescuing AID-tagged protein stability in HEK293T-TIR1 cells. Chronic depletion of AID-tagged ZNF143 and TEAD4 was rescued with wild-type ARF16-PB1 but not with ARF16-PB1-MT (Fig. 4C,D). However, we note that the ARF16-PB1 mutant protein is not as abundant as the wild type ARF16-PB1



**Figure 3.** ARF rescue promotes faster degradation kinetics upon addition of auxin. ZNF143 (A) and TEAD4 (B) are rapidly depleted upon auxin treatment. (C) A quantitative Western blot measuring the time-dependent degradation of ZNF143-AID in the presence of 500  $\mu$ M auxin. Note that ZNF143 is chronically depleted in these cells, and a longer exposure was needed to quantify auxin-induced reduction of ZNF143-AID levels. (D) A quantitative Western blot measuring 500  $\mu$ M auxin-induced ZNF143-AID degradation after ARF16-PB1 rescue. (E) We measured the intensity of ZNF143-AID bands from C and D using densitometry and fit the data using nonlinear regression and a one-phase decay equation.





**Figure 4.** The interaction of ARF-PB1 and AID mediates rescue of endogenous protein levels. (A) Positions indicated by the red arrows are highly conserved within the AUX-IAA family (Pfam family: PF02309). A seed alignment of the AUX-IAA family was generated using Pfam (El-Gebali et al. 2019), and the resultant FASTA files were visualized using WebLogo (Crooks et al. 2004). Positions on the X-axis are relative to ARF16 residue numbers. (B) The amino acids D994 and D998 of ARF16 and K114 of IAA17 are shown in the left panel. The amino acid side chains corresponding to D183 and D187 of IAA17 and K944 of ARF16 are highlighted in the right panel. Note that IAA17/ARF16 heterodimeric complexes are shown in both orientations and that these domains of IAA17 and ARF16 are within the same protein domain family. IAA17 and ARF16 sequences were modeled (Waterhouse et al. 2018) into each chain of an *A. thaliana* ARF5 homodimeric structure (Protein Data Bank [PDB] entry: 4CHK) (Nanao et al. 2014). These mutations in ARF16, which disrupt the electrostatic binding interface, fail to rescue chronic ZNF143 (C) and TEAD4 (D) degradation. (E) These mutations disrupt this coimmunoprecipitation of mCherry-AID with eGFP-tagged ARF16-PB1. Consistent with lower stability of the ARF16-MT in C and D, the mutant GFP-ARF16-PB1 plasmid was transfected at a concentration three times higher than the wild type to achieve comparable expression of each protein. (F) ARF16-PB1 is detected upon mCherry-AID immunoprecipitation; however, we were unable to detect TIR1 associating with AID.

(Fig. 4C,D), presumably because dimerization can stabilize exogenous ARF16-PB1. This result suggests that rescue of stability is mediated by the interaction between ARF16-PB1 and AID. To confirm a physical interaction, we performed a coimmunoprecipitation experiment. We found that NLS-mCherry-AID and eGFP-ARF16-PB1 interact in the absence of TIR1 (Fig. 4E). However, the ARF16 mutations are sufficient to disrupt the ARF16/AID interaction (Fig. 4E). The reciprocal immunoprecipitation confirmed the AID/ARF16-PB1 interaction, and we were unable to detect an interaction of TIR1 with AID (Fig. 4F). These data indicate that an ARF16-PB1/AID interaction mediates AID stability in the improved ARF-AID decon system.

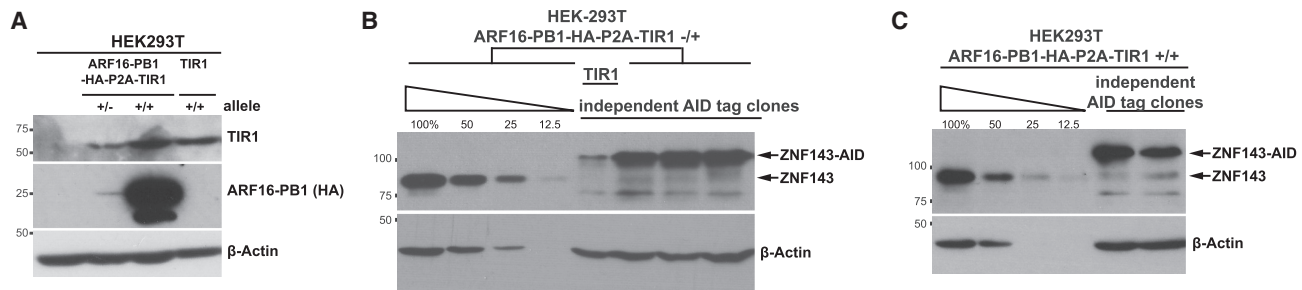
#### Expression of ARF-PB1 and TIR1 prior to AID-tagging preserves native protein levels

Rescuing chronic ZNF143 depletion does not result in full recovery of protein levels, and ~25% of the genes that are differentially expressed in the chronic depletion remain dysregulated upon ARF rescue (Fig. 2D). To further improve the system and preserve native levels of the tagged protein, we constitutively expressed ARF16-PB1 prior to

AID tagging. We generated a bicistronic construct with ARF-PB1 and TIR1 and an intervening P2A site, which separates the two proteins during translation. Genetic incorporation of this construct into the AAVS1 locus resulted in two independent clones that express ARF16-PB1 and TIR1 (Fig. 5A). ZNF143 AID tagging of both ARF-PB1/TIR1 HEK293T progenitor cell lines results in ZNF143 protein levels that are comparable with the parental lines (Fig. 5B,C).

#### ZNF143 activates transcription in cis

We directly and quantitatively assessed auxin-induced perturbation of ZNF143 by measuring genome-wide transcriptional changes upon auxin treatment. Conventional transcriptional profiling, such as RNA sequencing (RNA-seq), requires mature RNA to accumulate or degrade to detect changes in transcription. Delayed detection poses a challenge when measuring the immediate transcriptional response upon rapid protein depletion. To overcome this limitation, we measured nascent RNA using precision nuclear run-on sequencing (PRO-seq) (Core et al. 2008; Kwak et al. 2013; Mahat et al. 2016). PRO-seq quantifies the immediate effect that degradation has upon transcribing

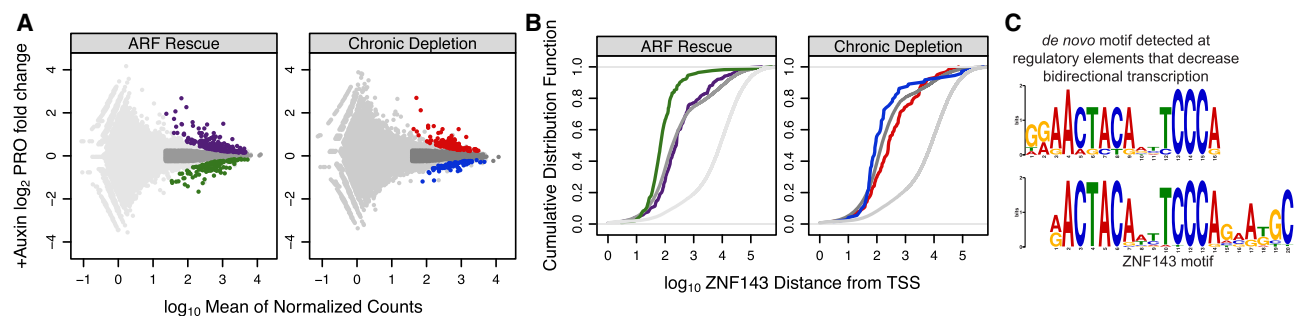


**Figure 5.** Expression of ARF-PB1 prior to AID tagging preserves ZNF143 protein levels. (A) We generated two progenitor cell lines that express HA-tagged ARF16-PB1 and TIR1 from the same promoter, separated by a P2A site. Unmodified HEK293T cells are within the first lane. The construct is incorporated into at least one allele of AAVS1 (second lane) and all AAVS1 alleles (third lane). Note that HEK293T cells are not strictly diploid. The fourth lane is the original TIR1-expressing progenitor cell line. (B) Three independent ZNF143-AID clones (sixth through eighth lanes) have protein levels comparable with those of the heterozygous ARF-PB1/TIR1 progenitor cells. Consistent with previous figures, TIR1 expression alone compromises ZNF143-AID protein levels. (C) Two independent ZNF143-AID clones (fifth and sixth lanes) derived from the homozygous ARF-PB1/TIR1 cells preserve ZNF143 protein levels.

RNA polymerases. ZNF143 was first characterized as a sequence-specific activator, as measured by reporter assays (Schuster et al. 1995). In order to identify direct ZNF143 target genes and primary response genes, we performed genome-wide nascent RNA profiling after 90 min of auxin treatment in the chronically depleted and ARF-rescued ZNF143-AID cell lines. We define “primary effect genes” as immediately regulated upon rapid ZNF143 depletion and “direct gene targets” as primary response genes that ZNF143 regulates in *cis*. Many genes are activated and repressed upon auxin treatment in both backgrounds (Fig. 6A). ZNF143 binding, as measured by ChIP-seq (chromatin immunoprecipitation [ChIP] combined with high-throughput sequencing) (The ENCODE Project Consortium 2012), is enriched proximal to the repressed gene class in the ARF rescue background (Kolmogorov-Smirnov two-sided  $P$ -value =  $1.1 \times 10^{-16}$ ) and not the activated gene class ( $P$ -value = 0.047) (Fig. 6B). In the chronic ZNF143 depletion background, the auxin-repressed gene class is not significantly closer to ZNF143-binding sites ( $P$ -value = 0.022), and genes within the activated gene class tend

to be further away from ZNF143-binding sites ( $P$ -value = 0.0017) (Fig. 6B). This supports recent genomic data, which found that many transcription factors (ER, GR, PPAR $\gamma$ , NF $\kappa$ B, and HSF) specialize to exclusively activate or repress transcription in *cis* (Carroll et al. 2006; Reddy et al. 2009; Guertin et al. 2014; Step et al. 2014; Schmidt et al. 2015, 2016; Duarte et al. 2016; Vockley et al. 2016).

Transcription factors can direct bidirectional transcription at their binding sites (Kim et al. 2010; Hah et al. 2013). Bidirectional transcription is a feature of gene promoters and enhancers (Core et al. 2014) and a measure of regulatory element activity. We tested whether we could detect changes in bidirectional nascent RNA production upon ZNF143 depletion. We implemented a discriminative regulatory element detection method (dREG) (Wang et al. 2019) to identify promoters and enhancers de novo. We performed differential bidirectional transcription analysis to identify regulatory elements that increase and decrease transcriptional activity upon auxin treatment (Supplemental Fig. S6). Next, we performed de novo motif analysis (Bailey et al. 2009) within



**Figure 6.** ZNF143 is a canonical transcriptional activator. (A) Genes are activated and repressed upon auxin treatment in both the chronic ZNF143-depleted background and the ARF rescue of the ZNF143 degradation background. Purple points are genes that increase expression, and green genes decrease expression upon auxin treatment in the ARF-rescued background. Dark-gray genes are expressed at levels comparable with those of activated and repressed genes but are confidently unchanged when auxin is added. (B) Cumulative distribution plots quantifying the relationship between the closest ZNF143 ChIP-seq peak and the start sites of genes within the regulated classes from A; trace colors correspond to categories in A. ZNF143-binding sites are closer to auxin-repressed genes only in the ARF rescue background, suggesting that ZNF143 functions to activate transcription in *cis*. (C) The ZNF143 motif was found de novo exclusively at regulatory elements that decrease bidirectional transcription upon auxin treatment.

the regulatory elements that increase or decrease bidirectional transcription. The canonical ZNF143 motif was found exclusively in the auxin-repressed regulatory elements (Fig. 6C). Taken together with the integrative ChIP-seq/PRO-seq analysis from Figure 6, A and B, we conclude that ZNF143 activates transcription of proximal genes and enhancers. These results serve as direct evidence that we are not only depleting ZNF143 protein levels but functionally perturbing ZNF143 activity using the ARF-AID system.

#### *ZNF143 targets are more responsive to perturbation upon ARF16 rescue*

Of the 168 genes that are classified as repressed upon auxin treatment in the ARF rescue background, 167 genes have a net negative change in auxin-induced gene expression in the chronic depletion background (Supplemental Fig. S7A). Upon auxin treatment, 87% (146 out of 168) of the genes have a greater magnitude of response in the ARF rescue background compared with chronic ZNF143-depleted cells (Supplemental Fig. S7A). These data show that the ARF rescue is more sensitive to detect auxin-induced changes in ZNF143-dependent transcription compared with the chronically depleted background. Seventy-eight percent (57 out of 73) of the auxin-repressed genes in the chronic ZNF143 depletion background are categorized as repressed in the rescue as well (Supplemental Fig. S7B); 70% (40 out of 57) are repressed to a greater magnitude in the rescue (Supplemental Fig. S7B). This analysis indicates that expressing ZNF143 at near-endogenous levels is necessary to detect a robust transcriptional response upon ZNF143 depletion.

#### *Auxin treatment activates the aryl hydrocarbon receptor (AHR) response*

In order to determine whether this system could be generally applied to study transcription factor function, we performed a control experiment to test whether auxin treatment alone affects transcription of human genes. Few genes are repressed upon auxin treatment (Supplemental Fig. S8A) regardless of FDR thresholds. However, over a range of FDR thresholds, we consistently observed that the activated genes (Supplemental Fig. S8) are enriched in AHR-binding sequences in their promoters and that the most enriched pathway for this gene set is the AHR pathway ( $q$ -value = 0.002) (Kuleshov et al. 2016). We found that AHR binding (Lo and Matthews 2012) is enriched proximal to the activated gene class ( $P$ -value =  $3.6 \times 10^{-11}$ ) and not the repressed gene class ( $P$ -value = 0.74) (Supplemental Fig. S8B,C). This control experiment highlights the importance of filtering AHR response genes from analyses when using any AID system to study transcriptional response.

#### *ZNF143 regulates paused RNA polymerase density*

Transcription can be regulated at various steps (Fuda et al. 2009; Scholes et al. 2017), including chromatin opening (Adelman et al. 2006; Guertin and Lis 2010; Morris et al.

2014), preinitiation complex formation/stability, and RNA polymerase II (Pol II) recruitment (Stargell and Struhl 1996), Pol II initiation (Sakurai and Fukasawa 2000; Govind et al. 2005; Esnault et al. 2008), Pol II pausing and release (Marshall and Price 1992; Adelman and Lis 2012), and elongation (Ardehali et al. 2009). General transcription machinery and cofactors directly catalyze these steps, but these factors are targeted to DNA by sequence-specific transcription factors, such as ZNF143. We sought to determine which transcription step(s) ZNF143 targets by characterizing the change in RNA Pol II profiles after rapid ZNF143 depletion (Supplemental Fig. S9). We found that the repressed gene class, which represents direct ZNF143 targets, shows dramatic changes in the pause region compared with the gene body (Fig. 7A,B; Supplemental Figs. S10, S11).

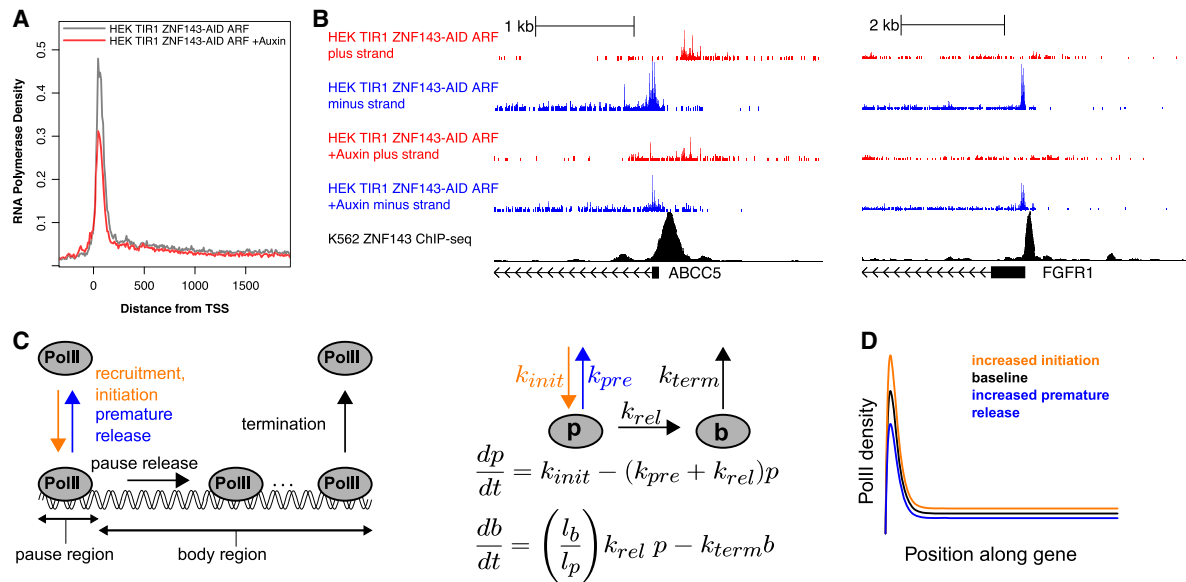
We implemented a mathematical modeling approach to better understand the potential mechanisms underlying decreased pause/body densities for repressed genes following ZNF143 depletion. We formulated a two-compartment model with dynamics for a pause region and gene body region. Model parameters included rate constants for transcriptional initiation/Pol II recruitment, premature pause release, pause release into productive transcription elongation, and termination of transcription (Fig. 7C). This model showed that only changes in initiation and premature pause release could account for a large magnitude of Pol II density change in the pause region relative to the gene body (Fig. 7D). These results suggest that decreases in the pause and gene body regions observed following ZNF143 depletion could be accounted for by either decreases in initiation or increases in nonproductive pause release. Moreover, the model predicted that the pause index would not change if initiation rate or premature pause release rate were affected by ZNF143 depletion, which is consistent with our result that the pause index at repressed genes is only modestly changed at genes that are repressed upon ZNF143 depletion (Supplemental Fig. S12). These combined experimental and modeling results suggest that ZNF143 regulates pausing density by facilitating RNA Pol II initiation or preventing premature dissociation of paused RNA Pol II.

## Discussion

Expression of ARF-PB1 improves the auxin-inducible degradation system. First, ARF-PB1 interacts with the AID-tagged protein to prevent degradation in the absence of auxin; thus, the tagged protein's abundance is more representative of native levels. Second, the rate of auxin-induced depletion is increased in the ARF-AID system. We demonstrated the power of the ARF-AID system by rapidly depleting the transcription factor ZNF143 and quantifying genome-wide changes in RNA polymerase density.

#### *Advantages of rapidly inducible degron systems*

Protein function can be studied by rapidly inducible degron systems in cases where translational fusion of the



**Figure 7.** ZNF143 regulates initiation/recruitment or pause stability. (A) The composite profile of Pol II density at all auxin-repressed genes indicates that Pol II pause density is compromised upon ZNF143 depletion. (B) Individual genes (ABCC5 and FGFR1) show compromised Pol II density in the pause region. (C) Model structure and key variables are highlighted in this schematic. A mathematical formulation of the two-compartment model, in which  $p$  refers to Pol II density at the pause region, and  $b$  refers to the density at the gene body region. (D) This plot represents a steady-state simulation for a reference model (black), a model in which transcriptional initiation was increased by 25% (orange), and a model in which premature pause release was increased by 60% (blue). The peak of the smooth curve was set to the steady-state pause level, and the plateau of the curve was set to the steady-state gene body level. Note that this plot captures the preferential effect on the pause region as compared with the gene body region.

degron tag does not disrupt protein function or protein stability. These systems provide advantages in interpretation of protein function because measurements taken immediately after protein dysregulation can be attributed directly to the protein of interest. In contrast, other techniques that are general, such as RNAi and genetic knockout, do not provide opportunities to assay phenotypes immediately after protein depletion due to the gradual or chronic nature of dysregulation. The newly developed dTAG system (Nabet et al. 2018) provides advantages comparable with those of ARF-AID, and exogenous expression of two additional proteins is not required. Endogenous tagging of BRD4 with FKBP12<sup>F36V</sup> dTAG did not result in dramatic protein depletion in the absence of the inducible degradation molecule dTAG-13 (Nabet et al. 2018). As the dTAG and ARF-AID systems become more widely adopted, we look forward to studies that systematically compare these different degron technologies. We found that auxin treatment alone causes undesired transcriptional changes at AHR target genes, but it is unclear whether dTAG-13 treatment results in off-target changes in cellular phenotypes. These types of side effects can be abrogated by including proper control experiments and depleting target proteins by multiple independent methods.

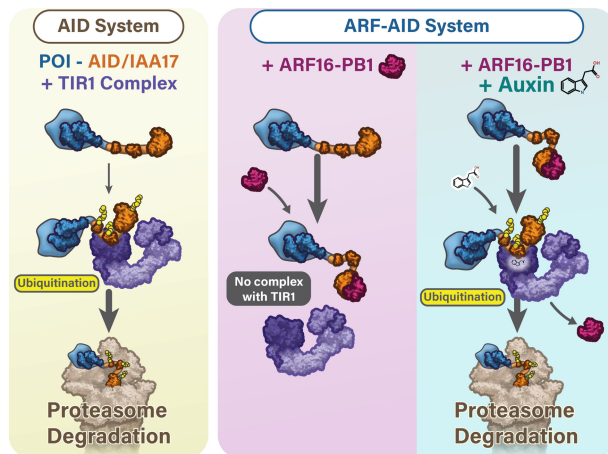
#### Possible mechanisms of ARF-mediated AID stabilization and rapid degradation

ARF transcription factors are a critical component of the plant auxin response system (Guilfoyle and Hagen 2007).

Here we found that ARF expression is an important component of engineered AID systems. The ARF-AID system confers two distinct advantages. (1) ARF expression limits auxin-independent degradation of target proteins, and (2) ARF expression promotes more rapid auxin-inducible degradation of AID-tagged proteins. ARF and Aux/IAA proteins harbor a conserved PB1 domain that can homodimerize or heterodimerize (Kim et al. 1997; Ulmasov et al. 1999). Mutations that interfere with ARF/AID interaction fail to rescue chronic auxin-independent AID degradation. TIR1 binds to domain II of Aux/IAA (Gray et al. 2001; Dharmasiri et al. 2003), and ARF binds to domains III and IV; therefore, ARF and TIR1 do not directly compete for the same binding surface of Aux/IAA. We propose that an ARF/AID interaction may cause conformational changes within AID that inhibit its interaction with auxin-unbound TIR1 (Fig. 8).

Auxin-inducible degradation of ZNF143 is more rapid in the ARF-AID system. There are many plausible and nonmutually exclusive possibilities for the observed difference in degradation kinetics. The simplest explanation is that promoting the stability of the AID-tagged protein increases its concentration, which directly affects degradation kinetics. Alternatively, ARF binding to AID may promote more efficient ubiquitination by exposing target residues to the auxin-bound TIR1 complex. ARF binding may cause allosteric changes in domain II of AID that promote a higher-affinity interaction with auxin-bound TIR1 compared with ARF-unbound AID. Recent work has shown that different Aux/IAA proteins can interact with





**Figure 8.** The ARF-AID system preserves protein of interest (POI) expression levels in the absence of auxin, and auxin treatment induces rapid degradation. (Left) The classical AID system can result in auxin-independent degradation of the AID-tagged proteins. (Right) In the ARF-AID system, ARF-PB1 binds to AID to prevent TIR1 association with AID, which prevents auxin-independent AID degradation by the ubiquitin-mediated proteasomal pathway. Auxin facilitates the interaction of TIR1 with AID and promotes dissociation of ARF and the subsequent ubiquitination and proteasome-mediated degradation of the AID-tagged protein.

TIR1 in the absence of auxin to influence the association of auxin with TIR1 (Calderón Villalobos et al. 2012). Although we did not detect an interaction of TIR1 and IAA17 in Figure 4, we cannot dismiss such an interaction in cells. Therefore, the ARF-AID system may promote faster kinetics because ARF binding affects AID structure, which in turn modulates the affinity of TIR1 and auxin. Future studies and systematic comparisons are needed to determine the mechanisms that contribute to a more rapid auxin-inducible response.

#### Molecular functions of ZNF143

Despite incredible advances in our understanding of the mechanisms of eukaryotic transcription and developments in systems biology, accurately predicting direct target genes and primary response genes of transcription factors remains a challenge. Proximal binding of a transcription factor to a gene is neither necessary nor sufficient to modulate gene expression. A fundamental question remains: How do transcription factors discriminate between genes in the genome in which to regulate? To begin to address this question, we must first define the set of genes regulated by the transcription factor of interest. Technical limitations preclude experimentally identifying a comprehensive set of primary response genes for the vast majority of transcription factors because we cannot rapidly induce or rapidly repress their activity. Perturbation methodologies that can be universally applied to any gene, such as RNAi, require days to efficiently deplete protein. This time frame of depletion poses a major barrier

to understanding transcription factor function because secondary (and beyond) effects dominate conventional depletion/knockout methods. Here, we show that applying the ARF-AID system to study transcription factor function overcomes these challenges.

The *Xenopus laevis* homolog of ZNF143, Staf (selenocysteine tRNA gene transcription activating factor), was first cloned and characterized nearly 25 yr ago (Schuster et al. 1995). This original report characterized the binding site of Staf within a regulatory element of the tRNA<sup>Sec</sup> gene and characterized the activator function of Staf using reporter assays (Schuster et al. 1995). Recent transcriptional profiling upon siRNA-mediated ZNF143 depletion identified many activated and repressed genes (Ngondo-Mbongo et al. 2013). ZNF143 depletion caused twice as many genes to decrease expression relative to the number of genes that increased expression. The investigators concluded that ZNF143 is primarily an activator but note that ZNF143 may be involved in repression. Our results corroborate the activation function of ZNF143 and indicate that although many genes are activated upon immediate ZNF143 depletion, ZNF143 does likely not act in *cis* to mediate repression. Importantly, we depleted ZNF143 for only 90 min and measured nascent RNA levels, so the repressive role of ZNF143 cannot be attributed to the postprimary response of ZNF143 dysregulation. Alternative mechanisms of immediate indirect repression, such as squelching (Guertin et al. 2014; Step et al. 2014; Schmidt et al. 2016), may be responsible for the observed repressive role of ZNF143. We further characterized ZNF143's role in activation and found that ZNF143 functions to control paused RNA Pol II density. Mathematical modeling indicates that ZNF143 either positively regulates RNA polymerase initiation or prevents nonproductive dissociation of paused Pol II.

ZNF143 is also involved in chromatin looping of distal enhancers to promoters (Bailey et al. 2015a). Enhancer-promoter looping frequently and preferentially occurs at promoters containing paused RNA Pol II (Ghavi-Helm et al. 2014). Therefore, we propose a model in which ZNF143 directly regulates the amount of paused Pol II on a given promoter, which facilitates enhancer looping.

Rapidly inducible systems, such as hormone signaling and heat-shock response, have contributed greatly to our understanding of transcriptional regulation. The success of these models in part is because the regulatory processes can be triggered instantaneously and tracked. New technologies and inhibitors that permit rapid and specific protein dysregulation promise to revolutionize the study of complex regulatory mechanisms.

#### Materials and methods

##### Cell lines

HEK293T cells were purchased from American Type Culture Collection (ATCC) and were grown in DMEM with 10% FBS, penicillin/streptomycin, and 5% glutamine. MCF7 cells were purchased from ATCC and were grown in DMEM with 10% FBS and penicillin/streptomycin. CENP-I-AID-eGFP DLD1-

OsTIR1 cells were generated in Ian Cheeseman's laboratory (McKinley and Cheeseman 2017) and grown in RPMI1640 medium with 10% FBS and penicillin/streptomycin.

#### Plasmids and constructs

OsTIR1 was integrated into AAVS1 locus of the HEK293T cells using the CMV-OsTIR1-PURO plasmid from Masato Kanemaki (pMK232; Addgene, 72834) (Natsume et al. 2016). OsTIR1 was integrated into the genome by CRISPR-Cas9-mediated repair using an sgRNA targeting AAVS1 safe harbor locus cloned into pSpCas9(BB)-2A-GFP from Feng Zhang (PX458; Addgene, 48138) (Mali et al. 2013; Ran et al. 2013; Natsume et al. 2016). The ARF16-PB1 domain (amino acids 878–1055) and ARF25 (MR and PB1 domain, amino acids 369–889, Os12t0613700-01) genes were codon-optimized for humans and synthesized from Bio Basic, Inc. We inserted ARF25-MR-PB1 into the eGFP-C2 vector. A nuclear localization signal (NLS) was added at the N terminus of the ARF16-PB1 domain, and NLS-ARF16-PB1 was inserted into eGFP-C2 vector digested with XhoI and HindIII by cold fusion cloning (pMGS36; Addgene, 126581). The CMV-OsTIR1-PURO plasmid was digested with AfeI, and we inserted the codon-optimized ARF16-PB1 domain separated by P2A from OsTIR1 to generate the ARF16-PB1-HA-P2A-OsTIR1 construct (pMGS46; Addgene, 126580). Additionally, we generated two GFP-ARF16-P2A-TIR1 plasmids that can be incorporated into a safe harbor locus: GFP-ARF16-PB1-MCS-P2A-OsTIR1 (pMGS55; Addgene, 129667) and GFP-ARF16-PB1-P2A-OsTIR1 (pMGS56; Addgene, 129668). The ARF16-PB1-HA-P2A-OsTIR1 plasmid was cotransfected with AAVS1 sgRNA (pMGS7; Addgene, 126582), and we selected puromycin-resistant clones to generate the homozygously integrated transgenic cell line. The resulting transgenic HEK293T-TIR1 and ARF16-PB1-HA-P2A-OsTIR1 cell lines were used to tag transcription factors with the AID tag using the CRISPR-Cas9 system. The NLS-mCherry-AID plasmid was constructed by digesting pCDNA5 vector with PmeI enzyme and inserting the NLS-mCherry-AID fragment using cold fusion cloning (System Biosciences).

#### Endogenous AID tagging in HEK293T cells

Endogenously AID-tagged TEAD4, ZNF143, and p53 cells were generated using CRISPR-mediated gene editing. sgRNAs that target the 3' end of the respective coding sequences were cloned into hSpCas9 plasmid (PX458; Addgene, plasmid 48138) (Ran et al. 2013). The linear donor was generated by PCR and gel-purified from a plasmid harboring a synthetic AID-P2A-hygromycin insert (pMGS54; Addgene, 126583). Note that this plasmid contains the full-length AID, and the mini-AID tag is unlikely to function in the ARF-AID system due to mini-AID's lack of domains III and IV. We amplified the AID-P2A-hygromycin insertion using primers that contain 50-nucleotide homology tails. The primers contained 5' phosphorothioate modifications to increase PCR product stability in the cell (Zheng et al. 2014). The primers used for making PCR donor fragments are reported in Supplemental Table S1. HEK293T cells were cotransfected with 1 µg of CRISPR/Cas9-sgRNA plasmid and 400 ng of linear donor PCR product using Lipofectamine 3000 in a six-well plate. Cells were expanded into 10-cm plates 2 d after transfection. The knock-in cells were selected by treating with 200 µg/mL hygromycin B 3 d after transfection. Individual clones were selected and confirmed by Western blotting and Sanger sequencing of PCR amplicons. The sgRNAs from Supplemental Table S2 were used for targeting the 3' end (C terminus of the protein) of the indicated genes.

#### EGFP-ARF16-PB1 stable cell lines

We transfected plasmids expressing NLS-ARF16-PB1 fused with eGFP at the C terminus or the eGFP-NLS-ARF16-PB1K944A, D994A, D998A mutant into each of the following: HEK293T, HEK293T-TIR1, and HEK293T-TIR1 cells in which either TEAD4 or ZNF143 were AID-tagged. Cells were expanded for 1 wk and GFP-sorted iteratively (three times) until we obtained a stable population of GFP-expressing cells.

#### Immunoprecipitation and immunoblotting

HEK293T or HEK293T-TIR1 cells were cotransfected with mCherry-NLS-AID and eGFP-ARF16-PB1 or eGFP-ARF16-PB1 mutant plasmids. The mutant eGFP-ARF16-PB1 plasmid was cotransfected at a concentration three times higher than the wild type to get comparable expression of the mutant protein. The coimmunoprecipitation data from Figure 4 were generated by separately transfecting mCherry-AID into HEK293T, HEK293T-eGFP-ARF16-PB1, and HEK293T-TIR1-eGFP-ARF16-PB1 cells. Cells were lysed 24 h after transfection in a buffer containing 50 mM Tris (pH 7.5), 150 mM NaCl, 0.5% NP-40, 1 mM EDTA, and protease and phosphatase inhibitors for 30 min on ice and then sonicated 30 sec on and 30 sec off for 15 cycles. Lysate was clarified by centrifugation at 12,000g for 15 min in 4°C. Anti-GFP antibody-conjugated magnetic beads (Chromotek, gtm-10) or anti-mCherry-conjugated affinity gel (Biolegend, 689502) were blocked with 1% BSA for 15 min. The beads were incubated with clarified lysate for 1.5 h at 4°C. The immunoprecipitate was recovered with DynaMag racks or by centrifuging at 3000g for 1 min. Beads were washed three times for 5 min in the lysis buffer, and 60 µL of 2× Laemmli buffer was added directly to the beads. The complex was heat-denatured for 5 min at 95°C. We used the following antibodies for the Western blots: anti-GFP (gift from Daniel Foltz, Northwestern University), anti-mCherry (rabbit, 1:5000; Abcam, ab183628), ZNF143 (1:5000; H00007702-MO1, Abnova), TEAD4 (1:1000; Santa Cruz Biotechnology, sc-101184), p53 (1:1000; Santa Cruz Biotechnology, DO1), anti-TIR1 (1:10,000; gift from Masato Kanemaki, Osaka University), β-Actin (1:5000; Sigma, A1978), β-Tubulin (AA2; gift from Todd Stukenberg, University of Virginia), and CENP-I (rabbit; gift from Todd Stukenberg, University of Virginia). Regression lines for kinetic data were fit using GraphPad Prism (Motulsky and Christopoulos 2004).

#### Drug treatment

Cells were treated with 10 µM MG132 for 4.5 h to test whether AID-tagged proteins were degraded through the proteasome pathway. A stock of 50 mM auxin was diluted to a final concentration of 500 µM auxin in the culture medium. The 50 mM stock was solubilized in DMSO for the 3-h auxin treatment PRO-seq experiments using the progenitor line. For all other experiments, 50 mM auxin was solubilized in water. A degradation rate upon auxin treatment was measured by treating cells with 10 µg/mL cycloheximide and 500 µM auxin and collecting samples at every 15 min for 4 h.

#### PRO-seq library preparation

Cell permeabilization was performed as described previously (Mahat et al. 2016). Cells were collected in 10 mL of ice-cold PBS after trypsinization and then collected and washed in 5 mL of buffer W (10 mM Tris-HCl at pH 7.5, 10 mM KCl, 150 mM sucrose, 5 mM MgCl<sub>2</sub>, 0.5 mM CaCl<sub>2</sub>, 0.5 mM DTT, 0.004

U/mL SUPERaseIN RNase inhibitor [Invitrogen], protease inhibitors [cOmplete, Roche]. The washed cells were then permeabilized with buffer P (10 mM Tris-HCl at pH 7.5, 10 mM KCl, 250 mM sucrose, 5 mM MgCl<sub>2</sub>, 1 mM EGTA, 0.05% Tween-20, 0.1% NP40, 0.5 mM DTT, 0.004 U/mL SUPERaseIN RNase inhibitor [Invitrogen], protease inhibitors [cOmplete, Roche]) for 3 min. Cells were washed again with 10 mL of buffer W before transferring into 1.5-mL tubes using wide-bore pipette tips. Finally, cells were resuspended in 500  $\mu$ L of buffer F (50 mM Tris-HCl at pH 8, 5 mM MgCl<sub>2</sub>, 0.1 mM EDTA, 50% glycerol, 0.5 mM DTT). After counting the nuclei, we generated 50- $\mu$ L aliquots with  $\sim 3 \times 10^5$  to  $5 \times 10^5$  cells that were snap-frozen in liquid nitrogen and stored at  $-80^\circ\text{C}$ . All centrifugations were done at 500g for 10-mL conical tubes and 2000g for 1.5-mL tubes at  $4^\circ\text{C}$ , and all buffers were maintained on ice. PRO-seq libraries were prepared as described previously (Duarte et al. 2016) with the following modifications. The libraries were amplified by PCR for a total of 10 cycles. We performed 5' decapping using RppH, 5' hydroxyl repair, 5' adapter ligation, and reverse transcription, while the 3' RNA biotin moiety was bound to magnetic streptavidin beads. We added an 8-base random unique molecular identifier (UMI) to the 5' end of the adapter that was ligated to the 3' end of the nascent RNA. We did not perform any size selection because we were willing to tolerate excessive adapter/adapter ligation products to ensure that our libraries were not biased against short nascent RNA insertions.

#### PRO-seq analyses

We removed adapters from the paired end 1 or single-end reads using CutAdapt (Martin 2011). Each 3' adapter harbored an 8-base UMI. We removed PCR duplicates based on the UMIs using fqdedup (<https://github.com/guertinlab/fqdedup>), trimmed UMIs with fastx\_trimmer ([https://github.com/agordon/fastx\\_toolkit](https://github.com/agordon/fastx_toolkit)), and implemented fastx\_reverse\_complement to generate the reverse complement sequence ([https://github.com/agordon/fastx\\_toolkit](https://github.com/agordon/fastx_toolkit)). We aligned reads to hg38 with Bowtie2 (Langmead et al. 2009), sorted aligned BAM files using SAMtools (Li et al. 2009), and used seqOutBias to generate bigWig files (Martins et al. 2018). We used the bigWig R package (<https://github.com/andremartins/bigWig>) and University of California at Santa Cruz (UCSC) Genome Browser utilities (Kent et al. 2010) to query bigWig files within genomic coordinates. Bedtools was used to parse genomic coordinate files and query for overlapping regions (Quinlan and Hall 2010). Differential nascent transcript abundance was measured by DESeq2 (Love et al. 2014). Bidirectional transcription was identified using dREG (Wang et al. 2019). MEME was used for de novo motif discovery within dREG-identified regulatory elements that change upon ZNF143 depletion; TOMTOM matched the ZNF143 motif (Bailey et al. 2015b) to a database that is curated by HOMER (Heinz et al. 2010). All of the analysis details and codes are available at [https://github.com/mjg54/znf143\\_pro\\_seq\\_analysis](https://github.com/mjg54/znf143_pro_seq_analysis). Raw sequencing files and processed bigWig files are available from Gene Expression Omnibus accession record GSE126919.

#### Model formulation

The dynamics for the concentrations or densities of RNA polymerases at pausing regions and gene bodies, defined as  $p$  and  $b$ , are described as follows:

$$\frac{dp}{dt} = k_{\text{init}} - (k_{\text{pre}} + k_{\text{rel}})p$$

and

$$\frac{db}{dt} = \left(\frac{l_p}{l_b}\right)k_{\text{rel}}p - k_{\text{term}}b,$$

where  $k_{\text{init}}$  is the rate of transcription initiation,  $k_{\text{pre}}$  is the rate constant for premature paused Pol II release,  $k_{\text{rel}}$  is the rate constant for the release of a paused Pol II as transcription proceeds into productive elongation in the gene body, and  $k_{\text{term}}$  is the rate constant for transcription termination. The term  $l_p/l_b$  is a ratio of the relative DNA segment lengths that is applied to adjust the gene body concentration based on the larger amount of DNA in the gene body as compared with the pause region. The term  $k_{\text{init}}$  implicitly accounts for the product of the unbound Pol II concentration and the rate constant for initiation. Concentration is considered as the number of RNA polymerases per length of DNA. Thus, this concentration is referred to as a density. We consider this model as though the units are dimensionless for the analysis of how specific rates influence the relative Pol II quantities at pause sites and within the gene bodies. The steady-state levels ( $p_{ss}, b_{ss}$ ) were found by setting  $dp/dt = db/dt = 0$ . Because understanding the effects of the relative pause region length was not our focus, we set  $r = l_p/l_b$  for the parameter sensitivity analyses:

$$p_{ss} = \frac{k_{\text{init}}}{k_{\text{pre}} + k_{\text{rel}}}$$

and

$$b_{ss} = \frac{rk_{\text{rel}}}{k_{\text{term}}}p_{ss} = \frac{rk_{\text{rel}}k_{\text{init}}}{k_{\text{term}}(k_{\text{pre}} + k_{\text{rel}})}.$$

The pause index is the relative density of reads in the pause region compared with the gene body ( $P_i = p_{ss}/b_{ss}$ ). The pause index is only dependent on  $k_{\text{term}}$  and  $k_{\text{rel}}$ , the rate constants for the termination of transcription and release into productive gene body elongation:

$$P_i = \frac{k_{\text{term}}}{rk_{\text{rel}}}.$$

#### Parameter sensitivity analysis

To examine the effects of each parameter on the pause and body concentrations, we considered high and low parameter values  $k^{(\text{hi})}$  and  $k^{(\text{lo})}$  and computed the changes in  $p_{ss}$  and  $b_{ss}$ . The following notation documents the change in  $p_{ss}$  when  $k$  is changed from a relatively high to a relatively low value:  $\Delta p(k) = p_{ss}[k^{(\text{hi})}] - p_{ss}[k^{(\text{lo})}]$ . First, we evaluated the effects of the transcription initiation rate ( $k_{\text{init}}$ ) on the steady state pause region and gene body Pol II concentrations:

$$\Delta p(k_{\text{init}}) = \frac{k_{\text{init}}^{(\text{hi})} - k_{\text{init}}^{(\text{lo})}}{k_{\text{pre}} + k_{\text{rel}}}$$

and

$$\Delta b(k_{\text{init}}) = \frac{rk_{\text{rel}}[k_{\text{init}}^{(\text{hi})} - k_{\text{init}}^{(\text{lo})}]}{k_{\text{term}}(k_{\text{pre}} + k_{\text{rel}})}.$$

Note that  $k^{(\text{hi})} > k^{(\text{lo})}$  by definition, so that  $\Delta p(k_{\text{init}}) > 0$  and  $\Delta b(k_{\text{init}}) > 0$ . Therefore, increasing the rate of transcription initiation will result in both pause region and gene body increases. We present results of single-parameter changes that can also be considered using a standard sensitivity analysis. For this example, the sensitivities were computed as follows:

$$\frac{\partial p_{ss}}{\partial k_{\text{init}}} = \frac{1}{k_{\text{pre}} + k_{\text{rel}}} > 0$$

and

$$\frac{\partial b_{ss}}{\partial k_{init}} = \frac{rk_{rel}}{k_{term}(k_{pre} + k_{rel})} > 0.$$

In general,  $\Delta p(k) \sim (\delta p / \delta k) [k^{(hi)} - k^{(lo)}]$  for small changes in  $k$ , and  $\text{sign}[\Delta p(k)] = \text{sign}(\delta p / \delta k)$ . Next, we considered the effects of varying the rate constant for premature pause release ( $k_{pre}$ ):

$$\Delta p(k_{pre}) = \frac{k_{init}[k_{pre}^{(lo)} - k_{pre}^{(hi)}]}{[k_{pre}^{(hi)} + k_{rel}][k_{pre}^{(lo)} + k_{rel}]}$$

and

$$\Delta b(k_{pre}) = \frac{rk_{rel}k_{init}[k_{pre}^{(lo)} - k_{pre}^{(hi)}]}{k_{term}[k_{pre}^{(hi)} + k_{rel}][k_{pre}^{(lo)} + k_{rel}]}.$$

These results show that increasing the rate constant for premature pause release will decrease both pause region and gene body Pol II concentrations [ $\Delta p(k_{pre}) < 0$  and  $\Delta b(k_{pre}) < 0$ ]. We next demonstrated that an increase in the rate constant for the release of Pol II into the gene body ( $k_{rel}$ ) decreases  $p_{ss}$  and increases  $b_{ss}$ :

$$\Delta p(k_{rel}) = \frac{k_{init}[k_{rel}^{(lo)} - k_{rel}^{(hi)}]}{[k_{pre} + k_{rel}^{(hi)}][k_{pre} + k_{rel}^{(lo)}]}.$$

and

$$\Delta b(k_{rel}) = \frac{rk_{init}k_{pre}[k_{rel}^{(hi)} - k_{rel}^{(lo)}]}{k_{term}[k_{pre} + k_{rel}^{(hi)}][k_{pre} + k_{rel}^{(lo)}]}.$$

It is interesting to note that if there is no premature pause release (i.e.,  $k_{pre} = 0$ ), the model predicts that a change in the rate of pause release into transcriptional elongation will not affect the concentration of RNA polymerases in the gene body if all other factors are identical [i.e.,  $\Delta b(k_{rel}) = 0$  for  $k_{pre} = 0$ ]. Finally, we evaluated the effect of modifying the rate constant for the termination of transcription:

$$\Delta p(k_{term}) = 0$$

and

$$\Delta b(k_{term}) = \frac{rk_{init}k_{pre}[k_{term}^{(lo)} - k_{term}^{(hi)}]}{k_{term}^{(hi)}k_{term}^{(lo)}(k_{pre} + k_{rel})}.$$

These results show that an increase in the rate constant for transcriptional termination does not affect the steady-state level of paused Pol II but will decrease the gene body concentration of RNA polymerases [ $\Delta p(k_{term}) = 0$  and  $\Delta b(k_{term}) < 0$ ]. The results from our parameter sensitivity analyses are illustrated in Table 1, where the entries for  $p_{ss}$  and  $b_{ss}$  indicate the effect of increasing each parameter.

**Table 1.** Parameter sensitivity summary

Parameter increased	Change in $p_{ss}$	Change in $b_{ss}$
$k_{init}$	Increase	Increase
$k_{pre}$	Decrease	Decrease
$k_{rel}$	Decrease	Increase
$k_{term}$	No change	Decrease

#### Relative effects of transcriptional initiation and premature release on Pol II distribution

Our preceding analyses show that consistent changes in the pause region and gene body (i.e., both either increase or decrease) are observed only following changes in the rate of transcriptional initiation or the rate constant for premature pause release ( $k_{init}$ ,  $k_{rel}$ ). Our experimental data show that repressed genes show decreases in both the pause region and the gene body region. These decreases in Pol II density were greater in the pause region in comparison with the gene body. Next, we documented the conditions under which the effects of varying the rates of initiation and premature release of the paused Pol II are greater for the pause region as compared with the gene body [e.g.,  $\Delta p(k_{init}) > \Delta b(k_{init})$ ]. Changes in transcriptional initiation are set as  $k_{init}^{(hi)} - k_{init}^{(lo)} = \Delta k$ :

$$\frac{\Delta k}{k_{pre} + k_{rel}} > \frac{rk_{rel}\Delta k}{k_{term}(k_{pre} + k_{rel})}.$$

This condition is satisfied for  $l_b/l_p > k_{rel}/k_{term}$  (recall  $r = l_p/l_b$ ). In general,  $l_b/l_p$  is a large value because the pause region (<100 base pairs [bp]) is much smaller than the gene body (approximately >10 kb). So,  $k_{rel}/k_{term} < l_b/l_p \sim 100$  must be obtained. For changes in premature release, the condition is as follows:

$$\frac{k_{init}(-\Delta k)}{[k_{pre}^{(hi)} + k_{rel}][k_{pre}^{(lo)} + k_{rel}]} > \frac{rk_{rel}k_{init}(-\Delta k)}{k_{term}[k_{pre}^{(hi)} + k_{rel}][k_{pre}^{(lo)} + k_{rel}]}.$$

This relation leads to the same constraint observed for  $k_{init}$ ,  $k_{rel}/k_{term} < l_b/l_p$ . Recall that the pause index is defined as  $P_i = k_{term}/rk_{rel} = (l_b/l_p)(k_{term}/k_{rel})$ . Therefore,  $P_i > 1$  for the same condition that constrains  $\Delta p(k_{pre}) > \Delta b(k_{pre})$  for changes in  $k_{init}$  and  $k_{pre}$ :  $k_{rel}/k_{term} < l_b/l_p$ . This demonstrates that for an arbitrary change in either the initiation rate or the premature pause release rate constant, the effect of the change in the pause region will be greater than the change in the gene body region whenever the pause region Pol II concentration is greater than that at the gene body. Furthermore, the effect ratio is identical for changes in  $k_{init}$  and  $k_{rel}$  and is equal to the value of the pause index:  $\Delta p/\Delta b = P_i$ .

#### Pause region and gene body model visualization

We aimed to generate plots in which steady-state levels of pause region and gene body concentration were imposed upon the peak and flat regions of a profile that is characterized by an exponential approach toward the peak followed by an exponential decay toward a stable plateau. We used a sum of exponential functions for the waveform:

$$\text{density}(\text{bp}) = \frac{pk_{\text{pause}}}{pk} \left\{ \frac{\text{bp}}{\tau} \exp\left[-\frac{(\text{bp} - \tau)}{\tau}\right] + pk_{\text{body}} \left[ 1 - \exp\left(-\frac{\text{bp}}{\tau}\right) \right] \right\},$$

where bp is the independent variable, and  $\tau$  is the exponential decay constant. The parameter  $pk$  is set to the root of the derivative as shown below so that  $pk_{\text{pause}}$  determines the peak of the waveform. The parameter  $pk_{\text{body}}$  is set such that the asymptotic gene body region decays to a desired level. The derivative of this waveform is

$$\frac{d}{dt} \text{density}(\text{bp}) = \frac{pk_{\text{pause}}}{pk} \left[ \frac{pk_{\text{body}}}{\tau} e^{-\text{bp}/\tau} - \frac{\text{bp}}{\tau^2} e^{(\tau - \text{bp})/\tau} + \frac{1}{\tau} e^{(\tau - \text{bp})/\tau} \right],$$

and the root of the derivative is given by the value of bp at the peak of the waveform [ $\text{max}[\text{density}]$ ]:

$$\text{bp} = \frac{\tau(pk_{\text{body}} + e)}{e}.$$

For  $\text{bp} \gg \tau$ , the gene body level asymptotically approaches  $pk_{\text{body}}/pk$ . We implicitly determined a value for  $pk_{\text{body}}$  that will



give a gene body level of a desired level. We selected a value of  $pk_{\text{body}}$  that will give a plateau  $pk_{\text{body}}pk_{\text{pause}}/pk$  of choice for a given setting of the pause peak  $pk_{\text{pause}}$ . After implicitly finding a value for  $pk_{\text{body}}$  that produces the gene body level of choice, the waveform is produced.

## Acknowledgments

We thank Arun Dutta, Dr. Anindya Dutta, Dr. Todd Stukenberg, and Dr. Prasad Trivedi for discussion and comments. Dr. Ryan Lewis from ScideLight (<http://scidelight.com>) illustrated Figures 1A and 8 and Supplemental Figures S1 and S3D. This work was funded by R35-GM128635 to M.J.G. and R21-HG009021 to L.C.

**Author contributions:** K.M.S. conceived the ARF-AID hypothesis. K.M.S. and B.D.M. performed the experiments. W.D.A. implemented the mathematical models. F.M.D., L.C., and M.J.G. developed the PRO-seq protocol. K.M.S. and M.J.G. conceptualized and developed the project. K.M.S. and M.J.G. designed the experiments. M.J.G. and K.M.S. analyzed the data. M.J.G. supervised the project and acquired the resources and funding. K.M.S., W.D.A., and M.J.G. wrote the original draft. K.M.S., B.D.M., W.D.A., F.M.D., L.C., and M.J.G. edited the manuscript.

## References

- Adelman K, Lis JT. 2012. Promoter-proximal pausing of RNA polymerase II: emerging roles in metazoans. *Nat Rev Genet* **13**: 720–731. doi:10.1038/nrg3293
- Adelman K, Wei W, Ardehali MB, Werner J, Zhu B, Reinberg D, Lis JT. 2006. *Drosophila* Paf1 modulates chromatin structure at actively transcribed genes. *Mol Cell Biol* **26**: 250–260. doi:10.1128/MCB.26.1.250-260.2006
- Ardehali MB, Yao J, Adelman K, Fuda NJ, Petesch SJ, Webb WW, Lis JT. 2009. Spt6 enhances the elongation rate of RNA polymerase II in vivo. *EMBO J* **28**: 1067–1077. doi:10.1038/emboj.2009.56
- Bailey TL, Boden M, Buske FA, Frith M, Grant CE, Clementi L, Ren J, Li WW, Noble WS. 2009. MEME suite: tools for motif discovery and searching. *Nucleic Acids Res* **37**: W202–W208. doi:10.1093/nar/gkp335
- Bailey SD, Zhang X, Desai K, Aid M, Corradin O, Cowper-Sal Lari R, Akhtar-Zaidi B, Scacheri PC, Haibe-Kains B, Lupien M. 2015a. ZNF143 provides sequence specificity to secure chromatin interactions at gene promoters. *Nat Commun* **6**: 6186. doi:10.1038/ncomms7186
- Bailey TL, Johnson J, Grant CE, Noble WS. 2015b. The MEME suite. *Nucleic Acids Res* **43**: W39–W49. doi:10.1093/nar/gkv416
- Calderon-Villalobos LI, Tan X, Zheng N, Estelle M. 2010. Auxin perception—structural insights. *Cold Spring Harb Perspect Biol* **2**: a005546. doi:10.1101/cshperspect.a005546
- Calderon Villalobos LIA, Lee S, De Oliveira C, Ivetac A, Brandt W, Armitage L, Sheard LB, Tan X, Parry G, Mao H, et al. 2012. A combinatorial TIR1/AFB-Aux/IAA co-receptor system for differential sensing of auxin. *Nat Chem Biol* **8**: 477–485. doi:10.1038/nchembio.926
- Cao S, Zhou K, Zhang Z, Luger K, Straight AF. 2018. Constitutive centromere-associated network contacts confer differential stability on CENP-A nucleosomes in vitro and in the cell. *Mol Biol Cell* **29**: 751–762. doi:10.1091/mbc.E17-10-0596
- Carroll JS, Meyer CA, Song J, Li W, Geistlinger TR, Eeckhoutte J, Brodsky AS, Keeton EK, Fertuck KC, Hall GF, et al. 2006. Genome-wide analysis of estrogen receptor binding sites. *Nat Genet* **38**: 1289–1297. doi:10.1038/ng1901
- Chapman EJ, Estelle M. 2009. Mechanism of auxin-regulated gene expression in plants. *Annu Rev Genet* **43**: 265–285. doi:10.1146/annurev-genet-102108-134148
- Core LJ, Waterfall JJ, Lis JT. 2008. Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. *Science* **322**: 1845–1848. doi:10.1126/science.1162228
- Core LJ, Martins AL, Danko CG, Waters CT, Siepel A, Lis JT. 2014. Analysis of nascent RNA identifies a unified architecture of initiation regions at mammalian promoters and enhancers. *Nat Genet* **46**: 1311–1320. doi:10.1038/ng.3142
- Crooks GE, Hon G, Chandonia JM, Brenner SE. 2004. WebLogo: a sequence logo generator. *Genome Res* **14**: 1188–1190. doi:10.1101/gr.849004
- Dharmasiri N, Dharmasiri S, Jones AM, Estelle M. 2003. Auxin action in a cell-free system. *Curr Biol* **13**: 1418–1422. doi:10.1016/S0960-9822(03)00536-0
- Dharmasiri N, Dharmasiri S, Estelle M. 2005. The F-box protein TIR1 is an auxin receptor. *Nature* **435**: 441–445. doi:10.1038/nature03543
- Duarte FM, Fuda NJ, Mahat DB, Core LJ, Guertin MJ, Lis JT. 2016. Transcription factors GAF and HSF act at distinct regulatory steps to modulate stress-induced gene activation. *Genes Dev* **30**: 1731–1746. doi:10.1101/gad.284430.116
- El-Gebali S, Mistry J, Bateman A, Eddy SR, Luciani A, Potter SC, Qureshi M, Richardson LJ, Salazar GA, Smart A, et al. 2019. The Pfam protein families database in 2019. *Nucleic Acids Res* **47**: D427–D432. doi:10.1093/nar/gky995
- The ENCODE Project Consortium. 2012. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**: 57–74. doi:10.1038/nature11247
- Esnault C, Ghavi-Helm Y, Brun S, Soutourina J, Van Berkum N, Boschiero C, Holstege F, Werner M. 2008. Mediator-dependent recruitment of TFIID modules in preinitiation complex. *Mol Cell* **31**: 337–346. doi:10.1016/j.molcel.2008.06.021
- Fuda NJ, Ardehali MB, Lis JT. 2009. Defining mechanisms that regulate RNA polymerase II transcription in vivo. *Nature* **461**: 186–192. doi:10.1038/nature08449
- Ghavi-Helm Y, Klein FA, Pakozdi T, Ciglar L, Noordermeer D, Huber W, Furlong EE. 2014. Enhancer loops appear stable during development and are associated with paused polymerase. *Nature* **512**: 96–100. doi:10.1038/nature13417
- Govind CK, Yoon S, Qiu H, Govind S, Hinnebusch AG. 2005. Simultaneous recruitment of coactivators by Gcn4p stimulates multiple steps of transcription in vivo. *Mol Cell Biol* **25**: 5626–5638. doi:10.1128/MCB.25.13.5626-5638.2005
- Gray WM, Kepinski S, Rouse D, Leyser O, Estelle M. 2001. Auxin regulates SCF<sup>TIR1</sup> dependent degradation of AUX/IAA proteins. *Nature* **414**: 271–276. doi:10.1038/35104500
- Guertin MJ, Lis JT. 2010. Chromatin landscape dictates HSF binding to target DNA elements. *PLoS Genet* **6**: e1001114. doi:10.1371/journal.pgen.1001114
- Guertin MJ, Zhang X, Conrod SA, Hager GL. 2014. Transient estrogen receptor binding and p300 redistribution support a squelching mechanism for estradiol-repressed genes. *Mol Endocrinol* **28**: 1522–1533. doi:10.1210/me.2014-1130
- Guilfoyle TJ, Hagen G. 2007. Auxin response factors. *Curr Opin Plant Biol* **10**: 453–460. doi:10.1016/j.pbi.2007.08.014
- Hah N, Murakami S, Nagari A, Danko CG, Kraus WL. 2013. Enhancer transcripts mark active estrogen receptor binding sites. *Genome Res* **23**: 1210–1223. doi:10.1101/gr.152306.112
- Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, Cheng JX, Murre C, Singh H, Glass CK. 2010. Simple combinations of

- lineage-determining transcription factors prime *cis*-regulatory elements required for macrophage and B cell identities. *Mol Cell* **38**: 576–589. doi:10.1016/j.molcel.2010.05.004
- Hoffmann S, Dumont M, Barra V, Ly P, Nechemia-Arbely Y, McMahon MA, Hervé S, Cleveland DW, Fachinetti D. 2016. CENP-A is dispensable for mitotic centromere function after initial centromere/kinetochore assembly. *Cell Rep* **17**: 2394–2404. doi:10.1016/j.celrep.2016.10.084
- Kent WJ, Zweig AS, Barber G, Hinrichs AS, Karolchik D. 2010. BigWig and BigBed: enabling browsing of large distributed datasets. *Bioinformatics* **26**: 2204–2207. doi:10.1093/bioinformatics/btq351
- Kepinski S, Leyser O. 2005. The *Arabidopsis* F-box protein TIR1 is an auxin receptor. *Nature* **435**: 446–451. doi:10.1038/nature03542
- Kim J, Harter K, Theologis A. 1997. Protein–protein interactions among the AUX/IAA proteins. *Proc Natl Acad Sci* **94**: 11786–11791. doi:10.1073/pnas.94.22.11786
- Kim TK, Hemberg M, Gray JM, Costa AM, Bear DM, Wu J, Harmin DA, Laptewicz M, BarbaraHaley K, Kuersten S, et al. 2010. Widespread transcription at neuronal activity-regulated enhancers. *Nature* **465**: 182–187. doi:10.1038/nature09033
- Korasick DA, Westfall CS, Lee SG, Nanao MH, Dumas R, Hagen G, Guilfoyle TJ, Jez JM, Strader LC. 2014. Molecular basis for auxin response factor protein interaction and the control of auxin response repression. *Proc Natl Acad Sci* **111**: 5427–5432. doi:10.1073/pnas.1400074111
- Kuleshov MV, Jones MR, Rouillard AD, Fernandez NF, Duan Q, Wang Z, Koplev S, Jenkins SL, Jagodnik KM, Lachmann A, et al. 2016. Enrichr: a comprehensive gene set enrichment analysis Web server 2016 update. *Nucleic Acids Res* **44**: W90–W97. doi:10.1093/nar/gkw377
- Kwak H, Fuda NJ, Core LJ, Lis JT. 2013. Precise maps of RNA polymerase reveal how promoters direct initiation and pausing. *Science* **339**: 950–953. doi:10.1126/science.1229386
- Langmead B, Trapnell C, Pop M, Salzberg S. 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* **10**: R25. doi:10.1186/gb-2009-10-3-r25
- Lavy M, Estelle M. 2016. Mechanisms of auxin signaling. *Development* **143**: 3226–3229. doi:10.1242/dev.131870
- Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, et al. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**: 2078–2079. doi:10.1093/bioinformatics/btp352
- Li SB, Xie ZZ, Hu CG, Zhang JZ. 2016. A review of Auxin response factors (ARFs) in plants. *Front Plant Sci* **7**: 47.
- Lo R, Matthews J. 2012. High-resolution genome-wide mapping of AHR and ARNT binding sites by ChIP-seq. *Toxicol Sci* **130**: 349–361. doi:10.1093/toxsci/kfs253
- Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**: 550. doi:10.1186/s13059-014-0550-8
- Mahat DB, Kwak H, Booth GT, Jonkers IH, Danko CG, Patel RK, Waters CT, Munson K, Core LJ, Lis JT. 2016. Base-pair-resolution genome-wide mapping of active RNA polymerases using precision nuclear run-on (pro-seq). *Nat Protoc* **11**: 1455–1476. doi:10.1038/nprot.2016.086
- Mali P, Yang L, Esvelt KM, Aach J, Guell M, DiCarlo JE, Norville JE, Church GM. 2013. RNA-guided human genome engineering via Cas9. *Science* **339**: 823–826. doi:10.1126/science.1232033
- Marshall NF, Price DH. 1992. Control of formation of two distinct classes of RNA polymerase II elongation complexes. *Mol Cell Biol* **12**: 2078–2090. doi:10.1128/MCB.12.5.2078
- Martin M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* **17**: 10. doi:10.14806/ej.17.1.200
- Martins AL, Walavalkar NM, Anderson WD, Zang C, Guertin MJ. 2018. Universal correction of enzymatic sequence bias reveals molecular signatures of protein/DNA interactions. *Nucleic Acids Res* **46**: e9. doi:10.1093/nar/gkx1053
- McKinley KL, Cheeseman IM. 2017. Large-scale analysis of CRISPR/Cas9 cell-cycle knockouts reveals the diversity of p53-dependent responses to cell-cycle defects. *Dev Cell* **40**: 405–420.e2. doi:10.1016/j.devcel.2017.01.012
- McKinley KL, Sekulic N, Guo LY, Tsinman T, Black BE, Cheeseman IM. 2015. The CENP-L-N complex forms a critical node in an integrated meshwork of interactions at the centromere-kinetochore interface. *Mol Cell* **60**: 886–898. doi:10.1016/j.molcel.2015.10.027
- Mendoza-Ochoa GI, Barrass JD, Terlouw BR, Maudlin IE, de Lucas S, Sani E, Aslanzadeh V, Reid JA, Beggs JD. 2019. A fast and tuneable auxin-inducible degron for depletion of target proteins in budding yeast. *Yeast* **36**: 75–81. doi:10.1002/yea.3362
- Morawska M, Ulrich HD. 2013. An expanded tool kit for the auxin-inducible degron system in budding yeast. *Yeast* **30**: 341–351. doi:10.1002/yea.2967
- Morris SA, Baek S, Sung MH, John S, Wiench M, Johnson TA, Schiltz RL, Hager GL. 2014. Overlapping chromatin-remodeling systems collaborate genome wide at dynamic chromatin transitions. *Nat Struct Mol Biol* **21**: 73–81. doi:10.1038/nsmb.2718
- Motulsky H, Christopoulos A. 2004. *Fitting models to biological data using linear and nonlinear regression: a practical guide to curve fitting*. Oxford University Press, New York.
- Nabet B, Roberts JM, Buckley DL, Paulk J, Dastjerdi S, Yang A, Leggett AL, Erb MA, Lawlor MA, Souza A, et al. 2018. The dTAG system for immediate and target-specific protein degradation. *Nat Chem Biol* **14**: 431–441. doi:10.1038/s41589-018-0021-8
- Nanao MH, Vinos-Poyo T, Brunoud G, Thévenon E, Mazzoleni M, Mast D, Lainé S, Wang S, Hagen G, Li H, et al. 2014. Structural basis for oligomerization of auxin transcriptional regulators. *Nat Commun* **5**: 3617. doi:10.1038/ncomms4617
- Natsume T, Kiyomitsu T, Saga Y, Kanemaki MT. 2016. Rapid protein depletion in human cells by auxin-inducible degron tagging with short homology donors. *Cell Rep* **15**: 210–218. doi:10.1016/j.celrep.2016.03.001
- Ngondo-Mbongo RP, Myslinski E, Aster JC, Carbon P. 2013. Modulation of gene expression via overlapping binding sites exerted by ZNF143, Notch1 and THAP11. *Nucleic Acids Res* **41**: 4000–4014. doi:10.1093/nar/gkt088
- Nishimura K, Fukagawa T. 2017. An efficient method to generate conditional knockout cell lines for essential genes by combination of auxin-inducible degron tag and CRISPR/Cas9. *Chromosome Res* **25**: 253–260. doi:10.1007/s10577-017-9559-7
- Nishimura K, Fukagawa T, Takisawa H, Kakimoto T, Kanemaki M. 2009. An auxin-based degron system for the rapid depletion of proteins in nonplant cells. *Nat Methods* **6**: 917–922. doi:10.1038/nmeth.1401
- Quinlan AR, Hall IM. 2010. Bedtools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**: 841–842. doi:10.1093/bioinformatics/btq033
- Ran FA, Hsu PD, Wright J, Agarwala V, Scott DA, Zhang F. 2013. Genome engineering using the CRISPR-Cas9 system. *Nat Protoc* **8**: 2281–2308. doi:10.1038/nprot.2013.143
- Reddy TE, Pauli F, Sprouse RO, Neff NF, Newberry KM, Garabedian MJ, Myers RM. 2009. Genomic determination of the

- glucocorticoid response reveals unexpected mechanisms of gene regulation. *Genome Res* **19**: 2163–2171. doi:10.1101/gr.097022.109
- Sakurai H, Fukasawa T. 2000. Functional connections between mediator components and general transcription factors of *Saccharomyces cerevisiae*. *J Biol Chem* **275**: 37251–37256. doi:10.1074/jbc.M004364200
- Samejima K, Ogawa H, Ageichik AV, Peterson KL, Kaufmann SH, Kanemaki MT, Earnshaw WC. 2014. A cell suicide module: auxin-induced rapid degradation of inhibitor of caspase-activated DNase (ICAD) induces apoptotic DNA fragmentation, caspase activation, and cell death. *J Biol Chem* **289**: 31617–31623. doi:10.1074/jbc.M114.583542
- Schmidt SF, Larsen BD, Loft A, Nielsen R, Madsen JGS, Mandrup S. 2015. Acute TNF-induced repression of cell identity genes is mediated by NF $\kappa$ B-directed redistribution of cofactors from super-enhancers. *Genome Res* **25**: 1281–1294. doi:10.1101/gr.188300.114
- Schmidt SF, Larsen BD, Loft A, Mandrup S. 2016. Cofactor squelching: artifact or fact? *Bioessays* **38**: 618–626. doi:10.1002/bies.201600034
- Scholes C, DePace AH, Sánchez Á. 2017. Combinatorial gene regulation through kinetic control of the transcription cycle. *Cell Syst* **4**: 97–108.e9. doi:10.1016/j.cels.2016.11.012
- Schuster C, Myslinski E, Krol A, Carbon P. 1995. Staf, a novel zinc finger protein that activates the RNA polymerase III promoter of the selenocysteine tRNA gene. *EMBO J* **14**: 3777–3787. doi:10.1002/j.1460-2075.1995.tb00047.x
- Shen C, Wang S, Bai Y, Wu Y, Zhang S, Chen M, Guilfoyle TJ, Wu P, Qi Y. 2010. Functional analysis of the structural domain of ARF proteins in rice (*Oryza sativa* L.). *J Exp Bot* **61**: 3971–3981. doi:10.1093/jxb/erq208
- Stargell LA, Struhl K. 1996. Mechanisms of transcriptional activation in vivo: two steps forward. *Trends Genet* **12**: 311–315. doi:10.1016/0168-9525(96)10028-7
- Step SE, Lim HW, Marinis JM, Prokesch A, Steger DJ, You SH, Won KJ, Lazar MA. 2014. Anti-diabetic rosiglitazone remodels the adipocyte transcriptome by redistributing transcription to PPAR $\gamma$ -driven enhancers. *Genes Dev* **28**: 1018–1028. doi:10.1101/gad.237628.114
- Szemenyei H, Hannon M, Long JA. 2008. Topless mediates auxin-dependent transcriptional repression during *Arabidopsis* embryogenesis. *Science* **319**: 1384–1386. doi:10.1126/science.1151461
- Tan X, Calderon-Villalobos LIA, Sharon M, Zheng C, Robinson CV, Estelle M, Zheng N. 2007. Mechanism of auxin perception by the TIR1 ubiquitin ligase. *Nature* **446**: 640–645. doi:10.1038/nature05731
- Tiwari SB, Hagen G, Guilfoyle TJ. 2004. Aux/IAA proteins contain a potent transcriptional repression domain. *Plant Cell* **16**: 533–543. doi:10.1105/tpc.017384
- Ulmasov T, Hagen G, Guilfoyle TJ. 1999. Dimerization and DNA binding of auxin response factors. *Plant J* **19**: 309–319. doi:10.1046/j.1365-3113.1999.00538.x
- Vockley CM, D'Ippolito AM, McDowell IC, Majoros WH, Safi A, Song L, Crawford GE, Reddy TE. 2016. Direct GR binding sites potentiate clusters of TF binding across the human genome. *Cell* **166**: 1269–1281.e19. doi:10.1016/j.cell.2016.07.049
- Wang D, Pei K, Fu Y, Sun Z, Li S, Liu H, Tang K, Han B, Tao Y. 2007. Genome-wide analysis of the auxin response factors (ARF) gene family in rice (*Oryza sativa*). *Gene* **394**: 13–24. doi:10.1016/j.gene.2007.01.006
- Wang Z, Chu T, Choate LA, Danko CG. 2019. Identification of regulatory elements from nascent transcription using dREG. *Genome Res* **29**: 293–303. doi:10.1101/gr.238279.118
- Waterhouse A, Bertoni M, Bienert S, Studer G, Tauriello G, Gumienny R, Heer FT, de Beer TAP, Rempfer C, Bordoli L, et al. 2018. SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Res* **46**: W296–W303. doi:10.1093/nar/gky427
- Zasadzińska E, Huang J, Bailey AO, Guo LY, Lee NS, Srivastava S, Wong KA, French BT, Black BE, Foltz DR. 2018. Inheritance of CENP-A nucleosomes during DNA replication requires HJURP. *Dev Cell* **47**: 348–362.e7. doi:10.1016/j.devcel.2018.09.003
- Zheng Q, Cai X, Tan MH, Schaffert S, Arnold CP, Gong X, Chen CZ, Huang S. 2014. Precise gene deletion and replacement using the CRISPR/Cas9 system in human cells. *Biotechniques* **57**: 115–124. doi:10.2144/000114196



## An improved auxin-inducible degron system preserves native protein levels and enables rapid and specific protein depletion

Kizhakke Mattada Sathyan, Brian D. McKenna, Warren D. Anderson, et al.

*Genes Dev.* published online August 29, 2019  
Access the most recent version at doi:[10.1101/gad.328237.119](https://doi.org/10.1101/gad.328237.119)

---

### Supplemental Material

<http://genesdev.cshlp.org/content/suppl/2019/08/22/gad.328237.119.DC1>

Published online August 29, 2019 in advance of the full issue.

### Creative Commons License

This article, published in *Genes & Development*, is available under a Creative Commons License (Attribution 4.0 International), as described at <http://creativecommons.org/licenses/by/4.0/>.

### Email Alerting Service

Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or [click here](#).

---

An advertisement for CRISPRmod. On the left is a colorful, abstract image of what appears to be a protein structure or cellular components in shades of red, orange, and purple. To the right, the text "Use CRISPRmod for targeted modulation of endogenous gene expression to validate siRNA data" is displayed in white on a dark blue background. In the bottom right corner is the "horizon" logo, which includes the word "horizon" in a sans-serif font and "a PerkinElmer company" in smaller text below it, all enclosed in a rounded rectangular box.

Use CRISPRmod for targeted modulation of endogenous gene expression to validate siRNA data

horizon  
a PerkinElmer company