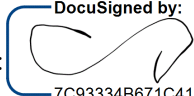**Distribution Agreement**

In presenting this thesis or dissertation as a partial fulfillment of the requirements for an advanced degree from Emory University, I hereby grant to Emory University and its agents the non-exclusive license to archive, make accessible, and display my thesis or dissertation in whole or in part in all forms of media, now or hereafter known, including display on the world wide web. I understand that I may select some access restrictions as part of the online submission of this thesis or dissertation. I retain all ownership rights to the copyright of the thesis or dissertation. I also retain the right to use in future works (such as articles or books) all or part of this thesis or dissertation.

Signature:

DocuSigned by:

7C93334B671C415...

Courtney Willett
Name

4/15/2024 | 11:45 AM EDT
Date

Title   Investigating the dynamics of transcription initiation and nucleosome turnover in Arabidopsis thaliana

Author   Courtney Willett

Degree   Doctor of Philosophy

Program   Biological and Biomedical Sciences

Genetics and Molecular Biology

Approved by the Committee

Roger Deal

Advisor

Karen Conneely

Committee Member

William Kelly

Committee Member

Judith Fridovich-Keil

Committee Member

Leila Rieder

Committee Member

Committee Member

Accepted by the Laney Graduate School:

_____

Kimberly Jacob Arriola, Ph.D, MPH
Dean, James T. Laney Graduate School

_____

Date

<u>Investigating the dynamics of transcription initiation and nucleosome turnover in</u>

<u>*Arabidopsis thaliana*</u>

by

Courtney Willett

B.S. Biology, Kennesaw State University, 2017

Advisor: Roger Deal, Ph.D.

An abstract of

A dissertation submitted to the Faculty of the

James T. Laney School of Graduate Studies of Emory University

in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Genetics and

Molecular Biology

2024

Doctoral Committee:

Dr. Karen Conneely

Dr. Judith Fridovich-Keil

Dr. William Kelly

Dr. Leila Rieder

Abstract

      The process of transcription is regulated by a variety of factors, such as RNA Polymerase II (RNA Pol II) initiation, histone post-translational modifications (PTMs) and cis-regulatory elements (CREs). Previous analysis of nascent RNA and chromatin patterns has revealed distinctions between plant and animal models. In plant models, nascent RNA transcripts and transcription-associated histone PTMs accumulate mainly in the gene body, suggesting that RNA Pol II does not initiate in the upstream direction of genes, as compared to animal models where RNA Pol II initiation can initiate bidirectionally. We wanted to investigate these patterns further at transcription start sites (TSSs) and at CREs to identify epigenetic and transcriptional differences between plant and animal models. Using a cross-species analysis of accessible chromatin, histone PTMs and nascent RNA data, we concluded that plants prefer strictly unidirectional RNA Pol II initiation both at the TSS and CREs, whereas *Drosophila* and *Homo sapiens* both show bidirectional transcriptional at either the TSS or CREs. Moreover, transcription can be highly disruptive to chromatin structure as the nucleosome must be at least partially disassembled to allow for RNA Polymerase II to transcribe the DNA. This disassembly and subsequent reassembly of the nucleosome can result in the loss of the original histone components and incorporation of non-canonical histone variants, which have a profound effect on epigenetic and transcriptional states. In particular, histone variant H3.3 is often incorporated into the nucleosome during transcription in a transcription-coupled manner and is associated with the maintenance of euchromatic, active regions of the genome. Knockdown of H3.3 in plant, yeast, and animal models predominantly affects environmental response genes; loss of H3.3 or its chaperone HIRA results in transcriptional defects but has a limited effects on global transcriptional outputs. Moreover, while is H3.3 incorporated across gene bodies in both plants

and animals, there is a distinct enrichment preferentially at the 3' end of genes in plants, further underscoring potential differences in nucleosome assembly between plants and animals. We developed an inducible system to measure incorporation of H3.3-GFP into chromatin. We observed a positive correlation between enrichment of H3.3-GFP incorporation and transcript levels, wherein the highest transcribed genes had the greatest H3.3-GFP incorporation at the 3' end of gene bodies over the span of 24 hours of induction, thus creating a useful system for measuring H3.3-GFP turnover dynamics relative to transcription activity. Overall, these studies highlight key differences in how plants regulate transcription and provide a novel tool for measuring nucleosome dynamics in Arabidopsis.

Investigating the dynamics of transcription initiation and nucleosome turnover in

Arabidopsis thaliana

by

Courtney Willett

B.S. Biology, Kennesaw State University, 2017

Advisor: Roger Deal, Ph.D.

A dissertation submitted to the Faculty of the

James T. Laney School of Graduate Studies of Emory University

in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Genetics and

Molecular Biology

2024

Doctoral Committee:

Dr. Karen Conneely

Dr. Judith Fridovich-Keil

Dr. William Kelly

Dr. Leila Rieder

scientist, you have fostered a home for my favorite people to grow into incredible scientists and people. From the bottom of my heart, thank you for everything.

I would also like to express extreme gratitude to my dissertation committee: Dr. Karen Conneely, Dr. Judith Fridovich-Keil, Dr. Bill Kelly, and Dr. Leila Reider. You all have provided invaluable feedback on my project throughout the years and also helped me to gain confidence in myself as a scientist. You all have gone above and beyond your roles as mentors and it means so much to me that you have approved me for this next chapter.

I owe my very presence in this program to my undergraduate mentor, Dr. Scott Nowak at Kennesaw State University. Until my senior year of college, I had written off a PhD. However, Scott encouraged me (threatened me with physical violence) to apply to Emory. He had enough confidence in me for the both of us and that made all the difference.

I would also like to extend my sincerest thanks to my friends outside of science. To my Dungeons & Dragons group (Bri, Ellen, Teelin and Dylan), thank you for creating a much-needed escape into fantasy and to our DM (Al), thank you for creating all new kinds of trauma to distract us from our jobs. To my community theater friends (Chloe, Molly, Rachel, David, Sam, Haley, and Ben), I am so grateful to have found you all and more importantly, defined myself outside of science in a creative safe space.

And finally, I cannot begin to express my sincerest thanks to my rock, my inspiration, my best friend, my loving partner and father of my cats, Al Such. I am continuously inspired by your work ethic and brilliance. While I was undergoing this six-year process, you managed to get a masters degree, land an outstanding job and get a promotion in an amazing field. You have taken on so much so that I could focus on this dissertation and I can only hope to support you as much as you support me. I love you so much and this project is dedicated to you.

Table of Contents

<div align="center">**Figures and Tables**</div>

Chapter 2

Figure 1: Histone modification enrichment and nascent transcripts across gene bodies in different organisms.

Figure 2: Histone modification enrichment and nascent transcript patterns around intergenic accessible chromatin regions (IARs).

Figure 3: Skew of GRO-seq signal across IARs.

Figure 4: Enrichment patterns around the TSS in multiple angiosperm species.

Supplemental Figure 1: Additional nascent RNA-seq datasets at TSSs and IARs in *Arabidopsis thaliana.*

Supplemental Figure 2: Comparison of nascent and steady-state RNA-seq at intergenic enhancer regions in Arabidopsis.

Supplementary Table 1: Publicly available dataset information

Supplementary Table 2: A. thaliana ChIP-seq antibody information

Supplementary Table 3: Data quality of *A. thaliana* root epidermal non-hair cell ChIP-seq datasets

Supplementary Table 4: Summary statistics for raw GRO-seq signal enrichment up and downstream of IARs for each species.

Chapter 3

Figure 1: H3.3-GFP is detectable after 5 hours of induction and strongest at 24 hours.

Figure 2: H3.3-GFP is turned over preferentially at the 3' end of gene bodies.

Figure 3: H3.3-GFP incorporation is positively correlated with gene expression level.

Figure 4: Estradiol-Induced H3.3-GFP incorporation at intergenic accessible regions.

Figure 5: Phosphate-starved plants have delayed H3.3-GFP incorporation at gene bodies.

Chapter 1

Introduction

Introduction

The basic unit of eukaryotic chromatin is the nucleosome. Comprised of an octamer of four core histone proteins (H2A, H2B, H3 and H4), the nucleosome plays a critical role in DNA compaction within the nucleus. The positioning of the nucleosome dictates access for DNA-binding proteins to functional locations within the genome. As such, chromatin is highly dynamic and responsive to developmental and environmental cues, specifically transcription.

Transcription initiation via RNA Polymerase II is the first step of transcriptional regulation in all eukaryotes. Cis-regulatory elements (CREs) define chromatin accessibility and allow for recruitment of transcriptional machinery to target specific genes in response to environmental and developmental cues. Histone variants can also provide information for transcriptional activity, such as histone variant H2A.Z. Previous studies in yeast found that H2A.Z enrichment at the +1 and -1 nucleosome of genes is indicative of bimodal transcription [1]. Similar H2A.Z patterns were observed in animal models [2, 3]. Plants, however, lack H2A.Z enrichment at the -1 nucleosome, suggesting strictly unidirectional transcription [4]. This is further supported by the patterns of histone post-translational modifications (PTMs) associated with RNA Polymerase II activity (e.g.. H3K4me3, H3K4me1) around the transcription start site (TSS). In animals and yeast, RNA Pol II initiates in both directions, regardless of the presence or absence of a gene on the opposite strand [5-8], whereas Arabidopsis shows unidirectional enrichment of these same PTMs at the TSS [9].

In chapter 1, we investigate this phenomenon by observing patterns of histone PTMs, H2A.Z positioning, chromatin accessibility, and nascent RNA transcripts at both CREs and TSSs. At a first glance, plant genomes are fundamentally different from animals. In plants, the majority of nongenic CREs fall within 2 kb of the TSS [10, 11] and fall preferentially upstream

of their target gene [12], whereas animals have been shown to have long-range enhancer interactions, both upstream and downstream from gene bodies [13-15]. Using an interspecies analysis of chromatin accessbility, nucleosome modifications and and nascent RNA (GRO/pNET/5-EU-RNA-seq), we concluded that bidirectional transcription occurs at the TSSs of *Drosophila* and *Homo sapiens*, but not in *Arabidopsis, Oryza sativa,* or *Glycine max*. This is consistent with previous studies that have analyzed transcription directionality (i.e. RNA Pol II initiation and histone PTMs) in plant, yeast and animal models [5-9]. We continued this analysis at CREs and observed a similar pattern of transcriptional unidirectionality in Arabidopsis, whereas *Drosophila* displayed both uni- and bidirectionality at CREs, further supporting that RNA Pol II initiation directionality is more tightly regulated in plants than in animals [16].

As mentioned earlier, processes like DNA replication and transcription disrupt the nucleosome structure. The majority of nucleosome assembly occurs during DNA replication, as the demand for core histones is doubled to accommodate the newly synthesized DNA. During S phase, cells express high levels of histones, in conjunction with existing ones, to reassemble chromatin during replication [17, 18]. Outside of S phase, there are several non-allelic sequence variants of histones, called histone variants, that are integrated into the nucleosome in a replication-independent manner. These histone variants alter the properties of the nucleosome they occupy and therefore alter epigenetic states of the surrounding chromatin landscape. This process of disassembly and reassembly of the nucleosome is referred to as turnover, as the histone proteins are ejected and replaced to quickly reform the nucleosome and maintain chromatin structural integrity. Histones are guided by histone chaperones, a diverse group of proteins that work to direct histone-DNA interactions [18, 19].

Of note, histone variant H3.3 is associated with euchromatic regions of the genome while H3.1 is associated with heterochromatic regions [20-26]. H3.3 differs from H3.1 at only four amino acids: A31T, F41Y, S87H and A90L [27]. H3.3 evolved independently in plants and animals, but is structurally similar; H3.3 differs from H3.1 at three of the four amino acids in both plants and animals, with plants having an additional substitution at residue 41 [27, 28]. The convergent evolution of H3.3 in plants and animals underscores its importance to maintaining structure and function of the eukaryotic genome.

Highly expressed genes have enrichment of H3.3 over the gene body, with plants showing a preference for this enrichment at the 3' end [26, 27, 29, 30]. Moreover, H3.3 deposition overlaps with active chromatin marks, such as H3K4me3 and RNA Pol II occupancy [25, 31]. However, H3.3 knockdown experiments in *Drosophila* [32], mouse embryonic stem cells [33, 34], and Arabidopsis [26] revealed limited changes in global transcription and H3K4me3 at the TSS; however, there were reduced transcription levels specifically in stress-response genes, suggesting H3.3's role in activation of genes during environmental stress [26]. There were also observed changes in DNA methylation, as well as H1 and H2A.Z occupancy over the gene body, indicating an antagonistic relationship between H3.3 and H1/H2A.Z [26]. Therefore, something about the presence of H3.3 inhibits repressive marks from encroaching on the gene body, further suggesting its role in maintaining euchromatin.

These results beg the question: what else about transcriptional regulation and H3.3 incorporation could potentially differ in plants versus animals? While H3.3 distribution patterns are similar in plants and animals, there are key differences in how H3.3 is incorporated into chromatin. In mammals, H3.3 is incorporated into genic regions via the histone chaperone protein HIRA (Histone Transcriptional Regulator A) and nongenic regions by ATRX/DAXX

(Alpha Thalassemia-mental Retardation X-linked syndrome/Death-domain Associated protein) [35, 36]. While Arabidopsis *hira atrx* double mutants have strong developmental defects, there are some interesting functional differences between plants and mammals [37]. Arabidopsis *atrx* mutants display loss of H3.3 in genic regions, counter to what is observed in mammals, suggesting functional divergence of ATRX [37]. Ergo, if the mechanisms in which H3.3 is deposited into chromatin differ between plants and animals, there could be differences in turnover dynamics of H3.3 as well.

At this point we understand that: 1) H3.3 shares functional similarity in maintaining transcriptionally active states between plants and animals, but evolved independently [28]; 2) transcription directionality is regulated differently in plants compared to animals [16]; and 3) H3.3 incorporation is mechanistically different in animals and plants[37]. In order to investigate the relationship between transcription activation and H3.3 in plants, we chose to investigate the incorporation of H3.3 into the nucleosome at transcriptionally active sites, thus providing insight into the role of nucleosome dynamics in maintaining active chromatin states.

Previous studies in yeast and mouse ESCs have measured rates of histone turnover and found that highly transcribed genes undergo greater rates of H3.3 turnover [29, 38]. This makes sense, considering the correlation between H3.3 deposition and transcription-associated marks (RNA Pol II, H3K4me3/H3K4me1, etc.) [26, 29, 39]. In chapter 2, we describe an inducible GFP-tagged H3.3 system in Arabidopsis that allows us to measure the rate of H3.3-GFP incorporation in real time. Similar to what has been observed in mice and yeast, highly transcribed genes have the greatest H3.3-GFP incorporation after 24 hours of H3.3-GFP induction. Moreover, we observe the same pattern at intergenic accessible chromatin regions (IARs), which serve as potential enhancer regions, consistent with what has been previously

described mouse embryonic stem cells [29]. Thus, we have created a useful system for measuring H3.3 incorporation and the dynamics of nucleosome turnover in plants.

We seek to understand the critical role histone variants play in shaping the chromatin landscape, specifically the preservation of epigenetic marks. A useful tool for studying epigenetic maintenance is environmental stress. Stress-induced transcriptional memory involves the activation of stress-related genes often accompanied by H3K4 hypermethylation, suggesting a model in which stress-induced epigenetic modulation extends the duration of active transcription [40, 41]. This begs the question: when transcription is activated at these sites, how does turnover of H3.3 affect the maintenance of these stress-induced epigenetic changes? Baurle et al [40] utilized a heat-shock (HS) inducible system to measure turnover of H3.3-GFP and found that histone turnover rates were lower at heat-shock (HS) memory genes, compared to non-memory genes; moreover, H3K4me3-marked histones were retained, suggesting that stress-induced hypermethylation is maintained by retention of the modified histones rather than replacement of existing H3 with new H3.3.

This relationship harkens back to previously described H3.3 knockdown experiments, which showed dysregulation of environmental response genes specifically, further supporting the relationship between H3.3 and transcription activation [26]. In the same vein, H3K4me3 is a marker of recent transcriptional activity (i.e. activation); ergo, the retention of H3.3, and as a result, H3K4me3, at response genes further indicates a role in turnover of H3.3 and gene activation [40].

Taken together, our results highlight fundamental differences and similarities in transcription activation and chromatin dynamics between plants and animals. During a cross-species meta-analysis of transcription-associated PTMs, accessible chromatin and nascent RNA

data, we observed that transcription in plants is highly regulated and strictly unidirectional at both the transcription start site and potential enhancer elements. In contrast, animals were observed to have bidirectional transcription at these sites. These stark differences in chromatin modifications and RNA Polymerase II activity suggest that there are potential differences in chromatin dynamics during transcription activation.

In addition, we developed an inducible system to measure the rate of GFP-tagged H3.3 incorporation genome-wide. Our data suggest a positive correlation between H3.3 incorporation and transcriptional activity, in keeping with what has been previously observed in animal and yeast models [42-44]. However, upstream gene body nucleosomes tend to retain existing H3 in plants, with a preference for turnover and H3.3 incorporation only near the transcription end sites. Overall, this system provides an incredibly useful foundation for investigating nucleosome dynamics in plants, with the ability to apply it to a variety of environmental conditions.

References

1.	Bagchi, D.N. and V.R. Iyer, *The Determinants of Directionality in Transcriptional Initiation.* Trends Genet, 2016. **32**(6): p. 322-333.
2.	Trinklein, N.D., et al., *An abundance of bidirectional promoters in the human genome.* Genome Res, 2004. **14**(1): p. 62-6.
3.	Heintzman, N.D., et al., *Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome.* Nat Genet, 2007. **39**(3): p. 311-8.
4.	Hetzel, J., et al., *Nascent RNA sequencing reveals distinct features in plant transcription.* Proc Natl Acad Sci U S A, 2016. **113**(43): p. 12316-12321.
5.	Core, L.J., J.J. Waterfall, and J.T. Lis, *Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters.* Science, 2008. **322**(5909): p. 1845-8.
6.	Neil, H., et al., *Widespread bidirectional promoters are the major source of cryptic transcripts in yeast.* Nature, 2009. **457**(7232): p. 1038-42.
7.	Seila, A.C., et al., *Divergent transcription: a new feature of active promoters.* Cell Cycle, 2009. **8**(16): p. 2557-64.
8.	Xu, Z., et al., *Bidirectional promoters generate pervasive transcription in yeast.* Nature, 2009. **457**(7232): p. 1033-7.
9.	Roudier, F., et al., *Integrative epigenomic mapping defines four main chromatin states in Arabidopsis.* Embo j, 2011. **30**(10): p. 1928-38.
10.	Maher, K.A., et al., *Profiling of Accessible Chromatin Regions across Multiple Plant Species and Cell Types Reveals Common Gene Regulatory Principles and New Control Modules.* Plant Cell, 2018. **30**(1): p. 15-36.
11.	Lu, Z., et al., *The prevalence, evolution and chromatin signatures of plant regulatory elements.* Nat Plants, 2019. **5**(12): p. 1250-1259.
12.	Jores, T., et al., *Identification of Plant Enhancers and Their Constituent Elements by STARR-seq in Tobacco Leaves.* Plant Cell, 2020. **32**(7): p. 2120-2131.
13.	Mifsud, B., et al., *Mapping long-range promoter contacts in human cells with high-resolution capture Hi-C.* Nat Genet, 2015. **47**(6): p. 598-606.
14.	Sanyal, A., et al., *The long-range interaction landscape of gene promoters.* Nature, 2012. **489**(7414): p. 109-13.
15.	Schoenfelder, S., et al., *The pluripotent regulatory circuitry connecting promoters to their long-range interacting elements.* Genome Res, 2015. **25**(4): p. 582-97.
16.	Silver, B.D., et al., *Differences in transcription initiation directionality underlie distinctions between plants and animals in chromatin modification patterns at genes and cis-regulatory elements.* G3 (Bethesda), 2024. **14**(3).
17.	Stewart-Morgan, K.R., N. Petryk, and A. Groth, *Chromatin replication and epigenetic cell memory.* Nat Cell Biol, 2020. **22**(4): p. 361-371.
18.	Robert, F. and C. Jeronimo, *Transcription-coupled nucleosome assembly.* Trends Biochem Sci, 2023. **48**(11): p. 978-992.
19.	Hammond, C.M., et al., *Histone chaperone networks shaping chromatin function.* Nat Rev Mol Cell Biol, 2017. **18**(3): p. 141-158.

20.   Chen, P., et al., *H3.3 actively marks enhancers and primes gene transcription via opening higher-ordered chromatin.* Genes Dev, 2013. **27**(19): p. 2109-24.

21.   Jin, C., et al., *H3.3/H2A.Z double variant-containing nucleosomes mark 'nucleosome-free regions' of active promoters and other regulatory regions.* Nat Genet, 2009. **41**(8): p. 941-5.

22.   Kraushaar, D.C., et al., *Genome-wide incorporation dynamics reveal distinct categories of turnover for the histone variant H3.3.* Genome Biol, 2013. **14**(10): p. R121.

23.   Mito, Y., J.G. Henikoff, and S. Henikoff, *Genome-scale profiling of histone H3.3 replacement patterns.* Nat Genet, 2005. **37**(10): p. 1090-7.

24.   Shu, H., et al., *Arabidopsis replacement histone variant H3.3 occupies promoters of regulated genes.* Genome Biol, 2014. **15**(4): p. R62.

25.   Stroud, H., et al., *Genome-wide analysis of histone H3.1 and H3.3 variants in Arabidopsis thaliana.* Proc Natl Acad Sci U S A, 2012. **109**(14): p. 5370-5.

26.   Wollmann, H., et al., *The histone H3 variant H3.3 regulates gene body DNA methylation in Arabidopsis thaliana.* Genome Biol, 2017. **18**(1): p. 94.

27.   Foroozani, M., D.H. Holder, and R.B. Deal, *Histone Variants in the Specialization of Plant Chromatin.* Annu Rev Plant Biol, 2022. **73**: p. 149-172.

28.   Waterborg, J.H., *Evolution of histone H3: emergence of variants and conservation of post-translational modification sites.* Biochem Cell Biol, 2012. **90**(1): p. 79-95.

29.   Deaton, A.M., et al., *Enhancer regions show high histone H3.3 turnover that changes during differentiation.* Elife, 2016. **5**.

30.   Zhao, F., et al., *The histone variant H3.3 promotes the active chromatin state to repress flowering in Arabidopsis.* Plant Physiol, 2021. **186**(4): p. 2051-2063.

31.   Wollmann, H., et al., *Dynamic deposition of histone variant H3.3 accompanies developmental remodeling of the Arabidopsis transcriptome.* PLoS Genet, 2012. **8**(5): p. e1002658.

32.   Hodl, M. and K. Basler, *Transcription in the absence of histone H3.3.* Curr Biol, 2009. **19**(14): p. 1221-6.

33.   Banaszynski, L.A., et al., *Hira-dependent histone H3.3 deposition facilitates PRC2 recruitment at developmental loci in ES cells.* Cell, 2013. **155**(1): p. 107-20.

34.   Yang, Y., et al., *HIRA complex presets transcriptional potential through coordinating depositions of the histone variants H3.3 and H2A.Z on the poised genes in mESCs.* Nucleic Acids Res, 2022. **50**(1): p. 191-206.

35.   Ricketts, M.D., et al., *Ubinuclein-1 confers histone H3.3-specific-binding by the HIRA histone chaperone complex.* Nat Commun, 2015. **6**: p. 7711.

36.   Lewis, P.W., et al., *Daxx is an H3.3-specific histone chaperone and cooperates with ATRX in replication-independent chromatin assembly at telomeres.* Proc Natl Acad Sci U S A, 2010. **107**(32): p. 14075-80.

37.   Duc, C., et al., *Arabidopsis ATRX Modulates H3.3 Occupancy and Fine-Tunes Gene Expression.* Plant Cell, 2017. **29**(7): p. 1773-1793.

38.   Rufiange, A., et al., *Genome-wide replication-independent histone H3 exchange occurs predominantly at promoters and implicates H3 K56 acetylation and Asf1.* Mol Cell, 2007. **27**(3): p. 393-405.

39.   Chen, P., et al., *Corrigendum: H3.3 actively marks enhancers and primes gene transcription via opening higher-ordered chromatin.* Genes Dev, 2021. **35**(9-10): p. 782.

40.   Pratx, L., et al., *Histone retention preserves epigenetic marks during heat stress-induced transcriptional memory in plants.* EMBO J, 2023. **42**(24): p. e113595.

41.   Oberkofler, V. and I. Baurle, *Inducible epigenome editing probes for the role of histone H3K4 methylation in Arabidopsis heat stress memory.* Plant Physiol, 2022. **189**(2): p. 703-714.

42.   Chory, E.J., et al., *Nucleosome Turnover Regulates Histone Methylation Patterns over the Genome.* Mol Cell, 2019. **73**(1): p. 61-72 e3.

43.   Dion, M.F., et al., *Dynamics of replication-independent histone turnover in budding yeast.* Science, 2007. **315**(5817): p. 1405-8.

44.   Li, Y., et al., *Replication-Independent Histone Turnover Underlines the Epigenetic Homeostasis in Adult Heart.* Circ Res, 2019. **125**(2): p. 198-208.

Chapter 2

Differences in transcription initiation directionality underlie distinctions between plants and animals in chromatin modification patterns at genes and cis-regulatory elements

[1]Department of Biology

[2]Graduate Program in Genetics and Molecular Biology

[3]Graduate Program in Biochemistry, Cell, and Developmental Biology

[#]Equal contributors

Chapter 2: Differences in transcription initiation directionality underlie distinctions between plants and animals in chromatin modification patterns at genes and cis-regulatory elements

Abstract

Transcriptional initiation is among the first regulated steps controlling eukaryotic gene expression. High-throughput profiling of fungal and animal genomes has revealed that RNA Polymerase II (Pol II) often initiates transcription in both directions at the promoter transcription start site (TSS), but generally only elongates productively into the gene body. Additionally, can Pol II initiate transcription in both directions at cis-regulatory elements (CREs) such as enhancers. These bidirectional Pol II initiation events can be observed directly with methods that capture nascent transcripts, and they are also revealed indirectly by the presence of transcription-associated histone modifications on both sides of the TSS or CRE. Previous studies have shown that transcription-associated histone modifications in the model plant *Arabidopsis thaliana* accumulate only in the gene body, suggesting that transcription does not initiate widely in the upstream direction in this plant. We compared transcription-associated histone modifications and nascent transcripts at both TSSs and CREs in *Arabidopsis thaliana*, *Drosophila melanogaster*, and *Homo sapiens*. Our results provide evidence for mostly unidirectional Pol II initiation at both promoters and CREs of *Arabidopsis thaliana*, whereas bidirectional transcription initiation is observed widely at promoters in both *Drosophila melanogaster* and *Homo sapiens,* as well as CREs in *Drosophila*. Furthermore, the distribution of transcription-associated histone modifications around TSSs in the *Oryza sativa* (rice) and *Glycine max* (soybean) genomes suggests that unidirectional transcription

initiation is the norm in these genomes as well. These results suggest that there are fundamental differences in transcriptional initiation between flowering plant and metazoan genomes.

Introduction

All organisms must control gene expression in a manner that is both cell type-specific and adaptive to changing cues. As such, transcription is a highly dynamic and regulated process, with many conserved mechanisms across eukaryotes. However, the nuanced differences in transcriptional regulation between eukaryotic kingdoms and even individual species are still being uncovered.

In an effort to further understand the relationships between transcription and chromatin, previous studies have analyzed correlations between transcription and histone post-translational modification (PTM) patterns. This led to the discovery of relationships between histone PTMs and different genomic regions, wherein unique PTM "chromatin signatures" can reflect whether a region is actively transcribed, poised, or constitutively silenced [1, 2]. As a pertinent example, H3K4me3 and H3K4me1 are an indirect result of RNA polymerase II (Pol II) activity. The Pol II-associated histone methyltransferase SET1 is responsible for the vast majority of H3K4 methylation in active genes in many eukaryotes, and its activity within the COMPASS complex leads to the distinct patterning of increased trimethylation at the TSS, dimethylation across the gene body, and monomethylation towards the 3' end [3-6]. In short, observing patterns of PTMs in conjunction with nascent transcript data such as Global Run-On Sequencing (GRO-seq)[7], Native Elongating Transcript sequencing (NET-seq) [8] or Nascent 5-EU-labeled RNA sequencing (Neu-seq) [9] can be a powerful tool for furthering our understanding of transcriptional activity and directionality.

It has been observed that in many animal genomes the distribution of histone PTMs indicative of active transcription is bimodal around the transcription start site (TSS), suggesting that transcriptional initiation is bidirectional in these eukaryotes [10]. Indeed, it has been demonstrated that Pol II often initiates in both directions at a given TSS in yeast and animals, regardless of the presence or absence of a gene on the opposite strand [7, 11-13]. In contrast, the plant *Arabidopsis thaliana* shows these same histone PTMs flanking just one side of the TSS, suggesting a unidirectional transcriptional mechanism [14]. This curious observation, which may indicate fundamental differences in the mechanism of Pol II initiation between plants and animals, prompted us to investigate transcriptional patterns at both TSSs and cis-regulatory elements (CREs) in different species.

Enhancers are CREs found in many organisms, including eukaryotes [15-18], bacteria [19] and viruses [20]. On a molecular scale, enhancer sequences are comprised of a modular collection of transcription factor binding motifs which act as an assembly platform for *trans*-acting factors [21, 22]. Sequence-specific transcription factors (TFs), general TFs, and co-factors associate with the enhancer and in turn recruit larger molecular machinery, including the Mediator complex, Pol II, nucleosome remodelers, and histone modifying proteins such as CPB/p300 to initiate transcription at the promoter [23]. Given that the binding of sequence-specific DNA binding proteins at enhancers displaces nucleosomes, these sites tend to be hypersensitive to nuclease digestion and can thus be identified at large through assays such as DNase-seq [24] [25] and ATAC-seq [26]. Additionally, studies in animal systems have identified characteristic histone PTMs associated with the nucleosomes that flank these CREs [27-31]. The set of histone PTMs associated with CREs in a variety of cell types and species include H3K4me1 and H3K4me3, which are deposited co-transcriptionally, in combination with H3K27ac or H3K27me3, depending

on the activity state of the enhancer. Recent evidence indicates that Pol II initiates at enhancer elements to generate enhancer RNAs (eRNAs) and this initiation, like at TSSs, is frequently bidirectional in animals [32, 33].

In this study, we integrated chromatin accessibility, ChIP-seq, and nascent transcript data from *Arabidopsis thaliana*, *Homo sapiens*, and *Drosophila melanogaster* to explore interspecies differences in chromatin modifications and transcriptional directionality. We first show evidence for mostly unidirectional transcription at TSSs in Arabidopsis and bidirectional transcription at those of human and Drosophila. We then examined CREs, defined as nuclease hypersensitive intergenic sites. Using the conserved set of enhancer histone PTMs H3K27ac, H3K27me3, H3K4me1, and H3K4me3, in conjunction with chromatin accessibility and nascent transcript data, we also find differences in transcriptional directionality at CREs between animal and plant genomes. While Drosophila shows frequent bimodal production of eRNAs and deposition of PTMs at accessible enhancer regions, Arabidopsis shows mainly unidirectional eRNA production with PTMs flanking only the transcribed side of the accessible chromatin region. Furthermore, analysis of ChIP-seq data from *Oryza sativa* (rice) and *Glycine max* (soybean) suggests that this pattern is not specific to *Arabidopsis* and may represent a fundamental difference in Pol II initiation processes between the plant and animal kingdoms. Taken together, our analyses provide important insight into the transcriptional dynamics of plants and suggest that differences in transcriptional directionality underlie the disparities observed in chromatin states between plant and animal epigenomes.

Results

Promoter Transcription Is Bidirectional In Animal Models And Preferentially Unidirectional In Arabidopsis

To address transcriptional directionality at the TSSs of protein coding genes, we combined nuclease hypersensitivity, ChIP-seq, and GRO-seq data from *Homo sapiens* (human), *Drosophila melanogaster*, and *Arabidopsis thaliana* (Supplementary Table 1). Publicly available data were used for human myeloid cells (all data), Drosophila S2 cells, (all data) and Arabidopsis root epidermal non-hair cells (ATAC-seq and ChIP-seq) and Arabidopsis seedlings (nascent transcript data). ChIP-seq data from Arabidopsis were generated from root epidermal non-hair cells in this study (Supplementary Tables 2 and 3). As much as possible, we attempted to analyze data from single cell types in order to minimize signal interference from different cell types. For the human analysis, this meant that DNase-seq and ChIP-seq are from CD34+ myeloid progenitor cells [34], while the GRO-seq analysis was from CD34+ myeloid progenitor cells cultured for 14 days and analyzed before terminal differentiation into erythrocytes [35]. For Arabidopsis, ATAC-seq [36] and ChIP-seq were from root epidermal non-hair cells but nascent transcriptome data were only available for seedlings [9, 37, 38] . Finally, Drosophila DNase-seq, ChIP-seq, and GRO-seq all came from S2 cells [39] [40]. This means that at a minimum, accessibility and histone PTM data are from the same cell type in each organism.

Gene-centric metaplots of the average ChIP-seq signal for H3K27ac, H3K27me3, H3K4me1, and H3K4me3 enrichment and chromatin accessibility across gene bodies in each of the three species are shown in **Figure 1A**. At this global scale, broad similarities are apparent in the pattern of chromatin accessibility relative to gene bodies. The region of maximum accessibility is restricted to a narrow peak 100-250 bp directly upstream of the TSS. Despite this fundamental similarity among organisms, a striking distinction emerges when the enrichments of H3K27ac,

H3K27me3, H3K4me1, and H3K4me3 are considered. The signal for these four histone modifications is clustered in a distinct bimodal pattern around the transcription start site (TSS) for both the human and Drosophila metaplots, with clear signals upstream and downstream of the TSS **(Figure 1A).** This pattern of enrichment is attributed to the bidirectional nature of animal promoters and their proclivity to produce transcripts from a single TSS in both the sense and antisense directions [3, 7, 41-43]. The elongating form of Pol II acts as a binding platform for histone modifying complexes, such as MLL3 and MLL4 in mammals, which deposit H3K4 methylation on the underlying histones successively through multiple rounds of elongation [44]. As such, the process of transcription itself is responsible for the surrounding deposition of this characteristic set of histone modifications [12]. This process leads to the enrichment of histone PTMs both upstream and downstream of the accessible TSS region in animals.
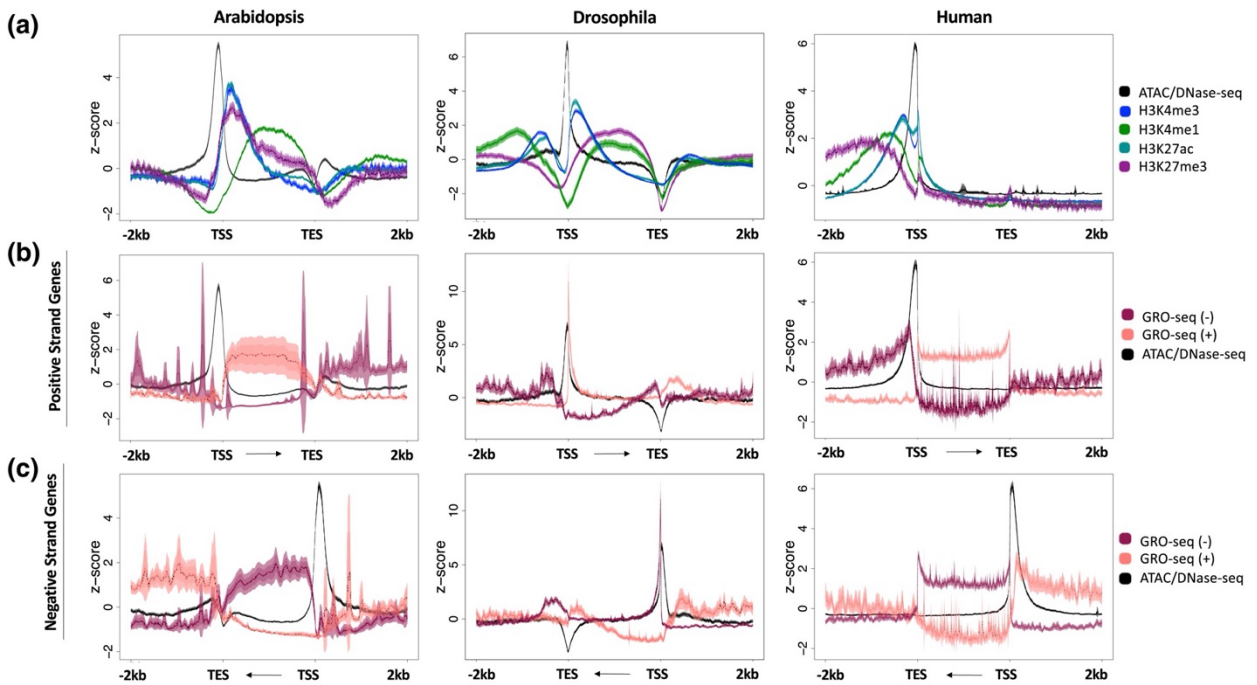
Figure 1: Histone modification enrichment and nascent transcripts across gene bodies in different organisms. a) Metaplots of average gene profiles of annotated Arabidopsis, Drosophila, and Human protein-coding genes. ChIP-seq signal for H3K27ac, H3K27me3, H3K4me1, and H3K4me3 are shown, along with chromatin accessibility data (ATAC-seq for Arabidopsis data; DNase-seq for Drosophila and Human data). b) Metaplots of the nascent transcriptional output (GRO-seq data) on Arabidopsis, Drosophila, and Human annotated positive strand and c) negative strand genes. Windows extend 2 kb upstream of the TSS and 2 kb downstream of the transcript end site (TES). For each dataset, several statistical metrics are shown. Solid lines represent the mean signal intensity (z-score normalized); the inner, dark-shaded region represents the standard error; and the outer, light-shaded region represents the 95% confidence interval (CI) of signal intensity.

In contrast, a unique pattern is seen at the TSS of the Arabidopsis metaplot. The histone modification ChIP-seq signals are most abundant at the 5' end of gene bodies, with the signal upstream of transcription start sites reduced to near background levels. Considering the mechanisms responsible for generating and maintaining the bimodal enrichment of histone modifications around animal TSSs, the absence of this pattern suggests that transcription in Arabidopsis may proceed nearly exclusively in the sense orientation, accounting for the sole downstream presence of histone marks. To examine this possibility more closely, we analyzed publicly available Global Run-On sequencing (GRO-seq) data from human, Drosophila, and Arabidopsis [38, {Gao, 2017 #506},[7]]. We separated all protein-coding genes across these genomes based by their strandedness, plotting GRO-seq signal at all plus strand genes (**Figure 1B**) and minus strand genes (**Figure 1C**) separately. Within gene bodies, the Arabidopsis metaplots

reveal that the directionality of the transcripts produced matches the directionality of the gene itself. In short, positive strand genes produce positive strand transcripts, while negative strand genes produce negative strand transcripts. Just upstream of the TSS in human and Drosophila genomes, transcripts running opposite of the genic direction are also produced, as is typical of bidirectional transcription at promoters [7, 45, 46] and enhancers [47-49]. This apparent absence of upstream signal in Arabidopsis indicates that transcription is strongly biased to be unidirectional in this organism. This pattern matches that observed in the enrichment of histone PTM signals, and further supports that histone PTM enrichment reflects transcriptional output. These findings were confirmed by analyzing two additional types of nascent transcript data at TSSs in Arabidopsis, NET-seq and 5-EU-RNA-seq [9, 37] (Supplementary Figure 2).

Intergenic CREs Are Unidirectionally Flanked By Characteristic Enhancer Marks In Arabidopsis

We next sought to investigate whether the pattern of transcriptional directionality observed at TSSs persists at cis-regulatory elements (CREs) in Arabidopsis. CREs have been shown to preferentially reside in regions of hyperaccessible chromatin [50, 51], and accessible sites have thus been used as markers of putative regulatory elements, such as enhancers [52]. To examine and compare gene-proximal CREs across species, we defined intergenic accessible regions (IARs) as nuclease hypersensitive sites that were outside of transcribed protein coding regions and were in the range of 100-2,000 bp away from a TSS to eliminate signals from protein-coding gene TSSs (Supplementary Figure 1). This strategy is supported by a number of previous findings. First, the majority of non-genic CREs in the *Arabidopsis*, rice, tomato, and *Medicago* genomes fall within 2 kb of the TSS [36] and many CREs have also been observed relatively close to the TSS in a

variety of other angiosperm species, both monocots and dicots [53]. Second, systematic analysis of enhancer element positioning by STARR-seq in plants suggested that enhancers preferentially work in the upstream position [54], thus the IARs are likely enriched for enhancer elements. To confirm that within our own data analysis, the GRO-seq signal within this window was driven by the presence of transient eRNAs, and not steady-state RNA transcripts from nearby TSSs, we clustered data from an RNA-seq experiment recently conducted in our lab onto the GRO-seq results. We found that generally, GRO-seq signal and RNA-seq signal were anti-correlated in these intergenic regions, supporting that this window was appropriate for identifying transient eRNAs while eliminating signal from TSSs (Supplementary Figure 1A). Additionally, IARs were located across the 100 bp to 2000 bp window (Supplementary Figure 1B). For the sake of consistency, we used this same window to select IARs when analyzing the Drosophila and human datasets.

We began by mapping four enhancer-associated histone modifications onto IARs. While regulatory elements are nucleosome-depleted regions where the frequent binding of *trans*-acting factors leaves the chromatin highly accessible, well-positioned nucleosomes flank the boundaries of these regions, often carrying characteristic histone modifications [55-57]. The metaplots of histone modification enrichment at IARs in all three species (**Figure 2A**) show symmetrical enrichment for H3K27ac, H3K4me1, and H3K4me3, which is in line with what has been previously reported in animal studies. Much like the pattern at gene bodies (**Figure 1**), H3K4me3 is enriched close to the accessible region, with H3K4me1 enrichment appearing more distally. These modifications are deposited during the process of transcription as the polymerase transitions from initiation to elongation [44], and are indicative of the production of eRNAs surrounding the accessible chromatin region (**Figure 2A**).

GRO-seq signal was also mapped over IARs (**Figure 2B**). The presence of GRO-seq signal at IARs further suggests that these potential enhancer regions are producing eRNAs. Previous studies in Arabidopsis and mammals have found that eRNAs have a role in gene regulation; interestingly, while eRNAs in mammalian cells are typically thought to be produced bidirectionally, eRNAs in Arabidopsis immunity-related CREs were shown to be mostly unidirectional [58, 59].
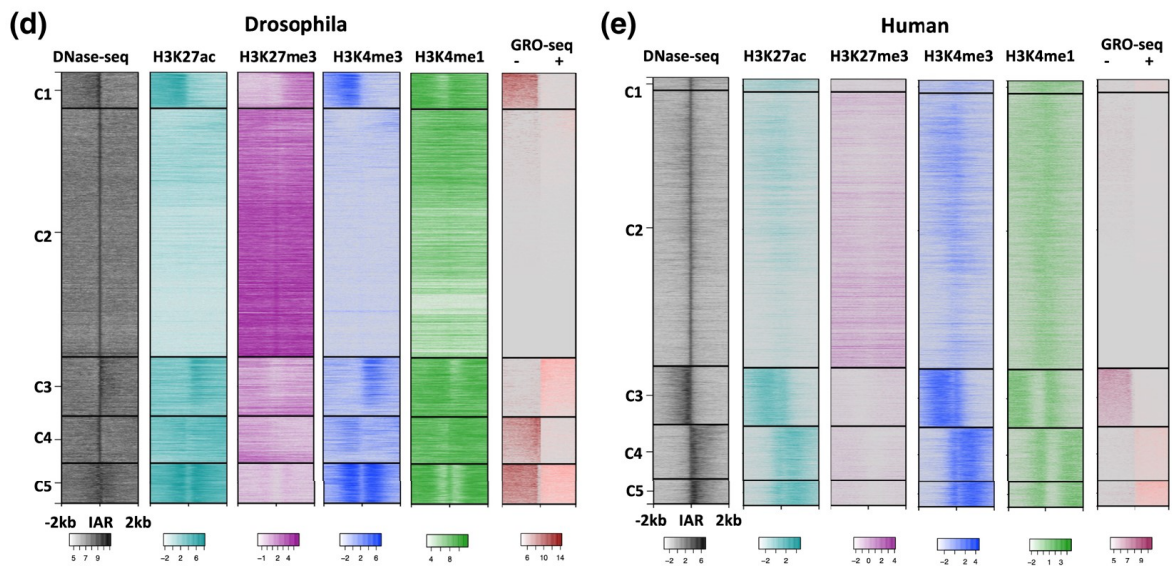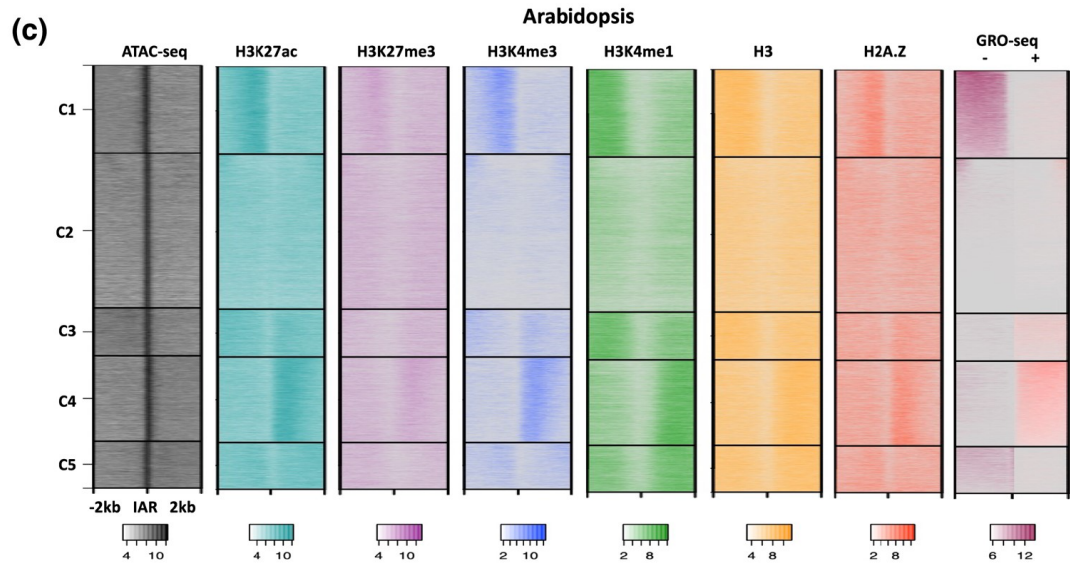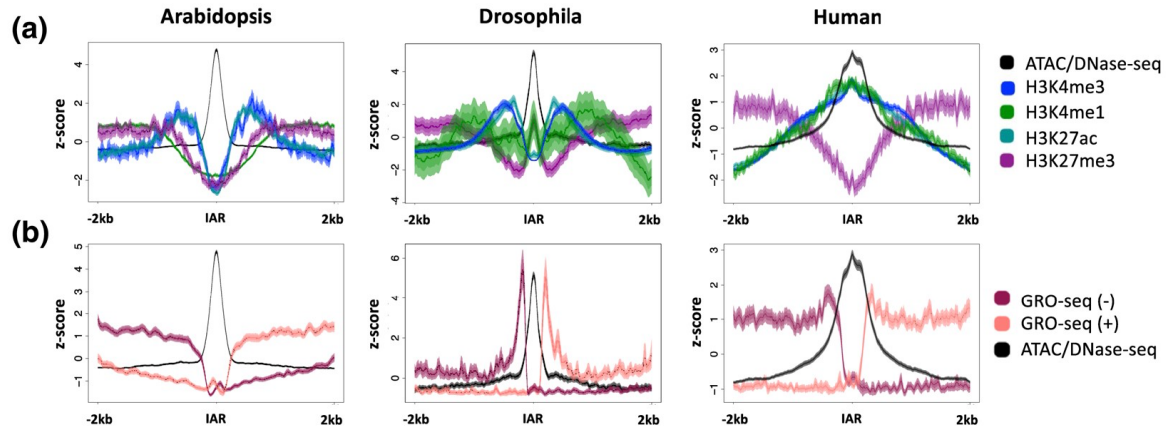
Figure 2: Histone modification enrichment and nascent transcript patterns around intergenic accessible chromatin regions (IARs). a) Metaplots of average histone modification and chromatin accessibility profiles at proximal IARs in Arabidopsis, Drosophila, and Human. b) Metaplot of chromatin accessibility and nascent transcriptional output (GRO-seq) at IARs. In both a and b, solid lines represent the mean signal intensity (z-score normalized); the inner, dark-shaded region represents the standard error; and the outer, light-shaded region represents the 95% confidence interval (CI) of signal intensity. c) Heatmaps of average profiles of proximal intergenic accessible regions in Arabidopsis (19,962), d) Drosophila (7,702 sites), and e) Human (12,664 sites). Heatmaps were log2 transformed and divided into five k-means clusters. Windows extend 2 kb upstream and 2 kb downstream of each accessible site.

While metaplots are useful in displaying the average signal across a group of loci, heatmaps expand on these trends by showing the precise signal pattern at each unique locus. Intergenic accessible chromatin regions across the *Arabidopsis*, Drosophila, and human genomes are displayed in a heatmap, grouped into subpopulations via k-means clustering in **Figure 2C-E**. In addition to H3K27ac, H3K27me3, H3K4me1, H3K4me3, and chromatin accessibility data, we also generated ChIP-seq datasets for histones H3 and H2A.Z in the *Arabidopsis* root epidermal non-hair cell type, which are also displayed in **Figure 2C**. H2A.Z is a histone variant associated with the flanking regions of active enhancers, while histone H3 is a core component of histone octamers and a marker of nucleosome occupancy [57]. High intensity signal can be seen in the center of the window in each ATAC-seq or DNase-seq heatmap, indicating a pronounced region of high chromatin accessibility. While there are clear differences between each of the species regarding the patterning of ChIP-seq and GRO-seq results, some universal patterns are observed.

Generally, active transcription as indicated by the presence of GRO-seq signal overlaps with H3K4me3, H3K4me1, and H3K27ac. This pattern is consistent with the presence of active enhancers, which are producing eRNAs. In each case H3K27ac and H3K4me3 are generally enriched closest to the center of the IAR, followed by H3K4me1 (**Figure 2C-E**). As has been documented previously in eukaryotic genes, lysine 4 of histone 3 is predominantly trimethylated at the 5' end of genes, with the modifications progressing to di- and monomethylation as transcriptional elongation proceeds [60, 61]. Finally, the Arabidopsis H2A.Z and H3 data show enrichments corresponding to higher GRO-seq signal, suggesting that these nucleosomes are particularly well positioned, promoting transcription. [55-57].

Distinct from most studied eukaryotes, however, the Arabidopsis chromatin shows dual enrichment for H3K27ac and H3K27me3 at the same loci **(Figure 2C, clusters 1 and 4).** This is in contrast to the pattern shown in **Figures 2D** and **2E**, which show more exclusivity of H3K27ac and H3K27me3 in Drosophila and human, respectively. While these modifications are considered to be mutually exclusive in animal models, this simultaneous enrichment has been documented in previous Arabidopsis chromatin studies [17]. Whether this is due to the presence of nucleosomes that are dually enriched with the marks – bearing one H3 with lysine 27 methylated, the other H3 lysine 27 acetylated – or due to different chromatin states among the genome copies in polyploid cells, is not yet clear.

Regarding transcription at IARs, the GRO-seq data suggest a clear preference for initiation in only one direction in Arabidopsis (**Figure 2C, Clusters 1,4**). The same unidirectional transcription pattern observed in clusters C1 and C4 in Arabidopsis was also observed in publicly available NET-seq and 5-EU RNA-seq [9, 37, {Szabo, 2020 #14], indicating a preference for unidirectional transcription at both TSSs and IARs (**Supplementary Fig 2**).

In contrast, *Drosophila* IARs display enrichment patterns consistent with either bidirectional (C5) or unidirectional (C1,3,4) transcription (**Figure 2D**). Many clusters show dual enrichment of H3K4me3 and H3K4me1, indicative of transcription (Clusters 1,3,4,5). In addition, many of the loci within these clusters also show moderate H3K27ac signal, as is typical of 'active' enhancers [62-64]. Other clusters show enrichment for H3K27me3/H3K4me1 characteristic of 'poised/inactive' enhancers (Cluster 2) [62, 63].

Similar to *Arabidopsis,* the human heatmaps exclusively show a preference for unidirectionality **(Figure 2E)**. GRO-seq enrichment is only seen on one side of the IAR, as mirrored by ChIP-seq signal (Cluster 3,4,5). This was somewhat surprising, as most literature supports bidirectional transcription of eRNAs in human cells [48, 65]. We can only speculate as to why the data we analyzed here are showing exclusively unidirectional enrichment. First, and perhaps most likely, is that the human gene-proximal CREs examined here behave differently than those at large. Alternatively, it could be that the cell type which we are analyzing does in fact preferentially produce eRNA transcripts in a unidirectional manner. Erythroblasts are the last stage before terminal differentiation, and during the final step to becoming a red blood cell, the nucleus is expelled. Just prior to this, the amount of RNA Polymerase II in the nucleus drops considerably, and it could be that with less overall transcriptional initiation at this stage, the production and/or ability to capture bidirectionally transcribed eRNAs decreases [66, 67]. This notion is consistent with reports that higher transcriptional activity correlates with increased production of bidirectional transcripts and eRNAs [68].

Finally, closer investigation into the relationships between nascent transcripts and histone PTMs reveals some notable differences between human and *Arabidopsis.* In each unidirectional cluster (**Figure 2E** clusters C3, 4, 5) in human, there is enrichment of H3K4me3 and H3K4me1

on both the transcribed and the non-transcribed side of the IAR, which is most reminiscent of the Drosophila bidirectional cluster (**Figure 2D**, cluster C5), however, the signal does not appear to stretch as far across the 2kb window in humans. It is possible that the ChIP-seq pattern present is a remaining hallmark of previous bidirectional transcription that started to diminish with lower rates of transcriptional activity. This is supported by a recent study that concluded that terminal erythroid maturation is associated with a loss of histone marks indicative of transcriptional elongation, but without a corresponding increase in heterochromatin marks [67].

In addition to examining transcriptional direction patterns qualitatively, we sought to quantify directionality. To calculate the bias of transcriptional signal across IARs observed in each genome, we calculated the average GRO-seq read depth signal both upstream and downstream for each IAR in each species. Statistical significance of differences between the two directions was determined using the Wilcoxon rank sum test, a non-parametric version of the two-sample t-test [69] and a Bonferroni-corrected significance threshold was applied [70] **(Figure 3)**. The apparent unidirectionality of Clusters C1 and C4 in Arabidopsis were supported by a statistically significant difference ($p < 2.2e\text{-}16$) between upstream and downstream signal (**Figure 3A)**. In contrast, Cluster C5 in Drosophila had an insignificant ($p = 0.01167$) difference between up and downstream signal, indicative of bidirectional enrichment of GRO-seq signal (**Figure 3B)**. As suggested by the qualitative analysis in Figure 2, human IARs preferentially showed unidirectional transcription, as indicated by significant differences in read density between the sides of each transcribed IAR (**Figure 3C)**.
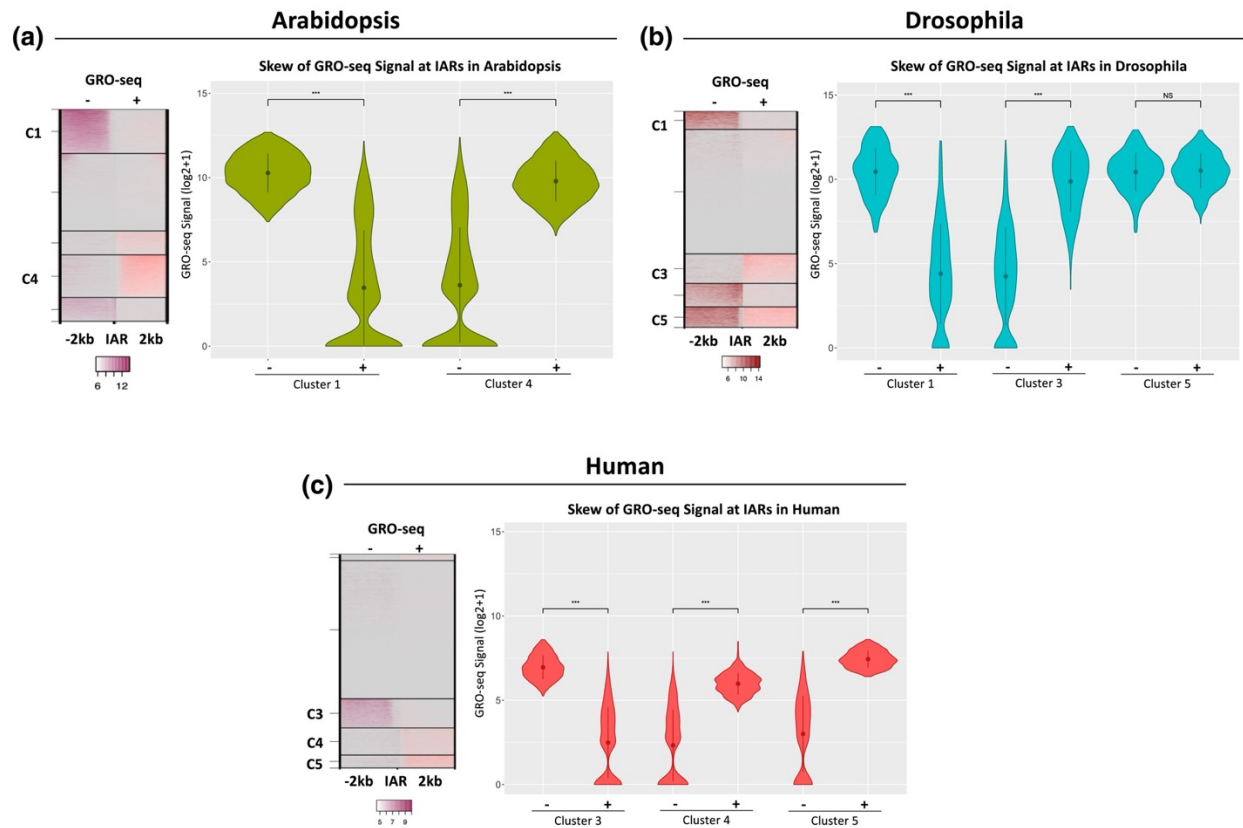
Figure 3: Skew of GRO-seq signal across IARs. GRO-seq signal was quantified on the upstream and downstream sides of each IAR to determine directionality of signal for a) Arabidopsis, b) Drosophila, and c) human. Directionality was calculated only for clusters that had strong GRO-seq signal on one or both sides of the IAR. Loci with low or no signal on either side were excluded from calculations. GRO-seq signal has been (log2 + 1) transformed. Statistical significance between up and downstream signal was calculated using a Wilcoxon Rank Sum test with a Bonferroni corrected significance threshold.

Analysis Of Other Flowering Plant Species Suggests A Similar Unidirectional Transcriptional Mechanism

In order to assess whether unidirectional transcription initiation is unique to Arabidopsis or a characteristic of flowering plants more generally, we analyzed publicly available chromatin accessibility and ChIP-seq data from *Oryza sativa* (rice) and *Glycine max* (soybean) [53]. We chose rice and soybean for these analyses to encapsulate both long evolutionary distances as well as different genome sizes. In our comparisons of rice, a monocot with a relatively small genome, and soybean, a dicot with a larger genome, we examined chromatin marks around the TSSs of plus and minus strand genes as a proxy for transcriptional directionality. Consistent with our observations in Arabidopsis, we observe one-sided flanking of transcription-associated histone modifications H3K4me1 and H3K4me3, as well as H2A.Z in the direction of the gene body (**Figure 4**). Thus, that these plants also show a pattern of PTMs consistent with unidirectional transcription at TSSs, suggesting that this may represent a general difference between metazoans and flowering plants.
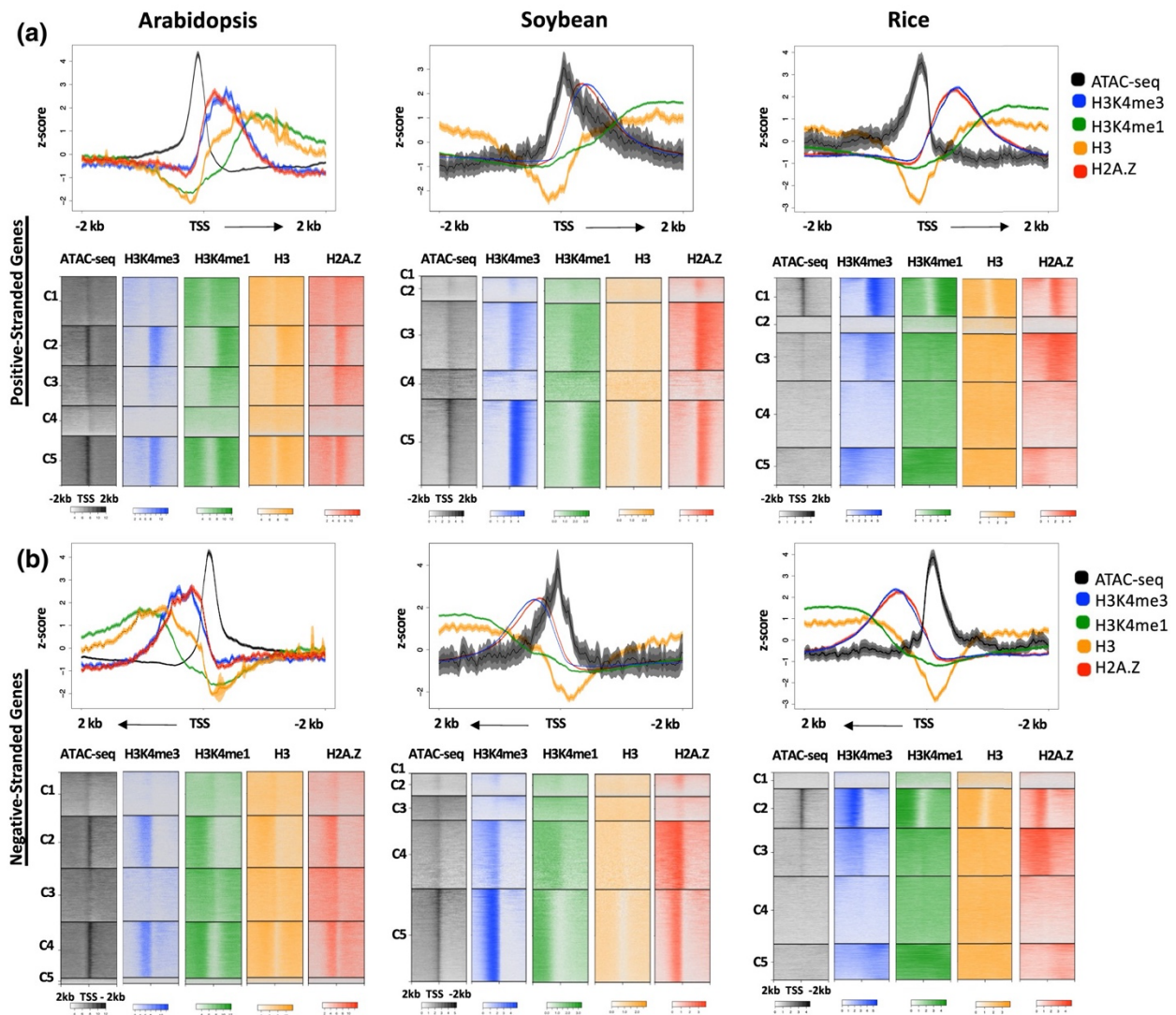
Figure 4: Enrichment patterns around the TSS in multiple angiosperm species. a) Metaplots and heatmaps of average H3K4me3, H3K4me1, H3 and H2A.Z signal and chromatin accessibility across positive-stranded gene TSSs in Arabidopsis (14,420 TSS sites), soybean (28,691 TSS sites), and rice (49,066 TSS sites). b) Metaplots and heatmaps of average H3K4me3 and H3K4me1, H3, and H2A.Z signal and chromatin accessibility across negative-stranded TSSs in Arabidopsis, soybean, and rice. Windows extend 2 kb upstream and 2 kb downstream of the TSS. In the metaplots, solid lines represent the mean signal intensity (z-score normalized); the inner, dark-

shaded region represents the standard error; and the outer, light-shaded region represents the 95% confidence interval (CI) of signal intensity. Heatmaps have been log2 transformed.

Discussion

In this study, we confirm the unique unidirectional histone PTM enrichment observed previously in *Arabidopsis* genes and show that this feature correlates with the direction of nascent transcription, a feature also present at intergenic CREs. Our study also provides ChIP-seq datasets with single-cell type resolution for the highly conserved set of histone PTMs H3K27ac, H3K27me3, H3K4me1, and H3K4me3, as well as histone variant H2A.Z in the model plant *Arabidopsis thaliana*. When examined alongside ATAC-seq data previously generated by our lab and available GRO-seq, NET-seq and 5-EU RNA-seq data, stark differences become apparent between plant and animal species. Histone PTM enrichment at TSSs of protein-coding genes in human and Drosophila show a distinct bimodal distribution pattern around the TSS, while the enrichment pattern found around TSSs in Arabidopsis and two other plant species is noticeably missing signal upstream of the TSS (**Figures 1 and 4**). This bimodal pattern has been shown to be indicative of bidirectional transcription at the promoter [7, 45, 46], strongly implying that in Arabidopsis, rice, and soybean, transcription is more tightly regulated at the level of initiation.

Intergenic CREs in animal genomes correspond with the hallmark enhancer histone PTM pattern, being flanked by histones with high levels H3K27ac (active) or H3K27me3 (poised) and H3K4me1, with relatively less H3K4me3 **(Figure 2).** While metaplots of PTMs at intergenic accessible sites suggest a similar bimodal pattern in Arabidopsis, clustered heatmaps revealed that histone PTMs flank the region either upstream or downstream of the accessible site. Additionally,

H3K27me3 and H3K27ac are not mutually exclusive, as they are in animal species. Taken together, these findings indicate that the same symmetrical PTMs used in animals to identify enhancer regions will not apply to plant species, where these marks generally flank only the single transcribed side of the element. Finally, nascent transcript data reveal that putative eRNAs at Arabidopsis CREs are produced unidirectionally (**Figure 2C**), consistent with recent findings in Arabidopsis immunity-related CREs [59]. What the difference is in gene structure, chromatin landscape and/or promoter:enhancer interactions between those enhancers that produce unidirectional eRNAs and those that are bidirectional is a rich area for future study.

Overall, the results of this investigation indicate that genuine differences exist between the plant and animal kingdom at the level of transcriptional initiation. While the elongation of protein-coding transcripts in the sense direction appears to be preferred across all eukaryotes, the results of this study suggest that this direction is preferred with near exclusivity in transcriptional initiation at TSSs in plants, while animal transcription initiation is more promiscuous.

These results beg the question, what is the reason for this stark contrast between transcriptional directionality in Arabidopsis and these animal species? One potential reason is that, distinct from animals, plants contain RNA-directed DNA methylation pathways that can silence portions of the genome, and these pathways are primarily targeted by siRNAs, which could be generated by overlapping divergent transcripts. This suggests that if allowed to proceed unchecked, the production of reverse transcripts could disrupt the plant epigenome, resulting in a strong selective pressure to keep the generation of bidirectional eRNAs tightly regulated, if not eliminated entirely.

How, then, are plants able to prevent the production of reverse strand transcripts? Recent findings from Hi-C data with single-gene resolution may shed some light on the question. Rather

than forming the large topologically associated domains (TADs) found in mammals, the *Arabidopsis* genome is preferentially organized into small, local gene loops, where the 5' and 3' ends of a gene directly interact [71]. The constrained geometry of these gene loops has been suggested to eliminate bidirectional transcription, forcing Pol II to transcribe in the sense direction alone [72]. This organizational scheme and relative lack of long-range interactions could explain why the majority of intergenic accessible chromatin sites we identified were preferentially located proximally upstream of their nearest gene [36]. The Pol II-associated factor Ssu72 is responsible for maintaining the association between the gene ends in yeast; when this factor is mutated, the gene loop structure is abolished and bidirectional transcripts are produced [72, 73]. While it is not yet known whether plants contain a functional ortholog of Ssu72, these findings suggest that analogous differences in higher order chromatin structure may be responsible for the observed distinctions in transcriptional direction.

Additionally, the presence or absence of 5' Pol II pausing in *Arabidopsis* has been contested. While some studies support the absence of pausing in both *Arabidopsis* and maize [38], another more recent study utilizing NET-seq has shown evidence to the contrary [37]. It is worth noting even in that study, 5' pausing does not appear to be as tightly regulated in Arabidopsis as it is in humans. These findings, in conjunction with the lack of negative elongation factor (NELF) in plants [74] further support that transcriptional regulation in plants has notable distinctions from the mechanisms in animals and is likely more heavily regulated at the level of initiation.

In the future, it will be interesting to parse out how particular chromatin features observed here, such as the presence or absence of specific histone PTMs or variants, may contribute to the regulation of initiation. It is likely that many PTMs and variants are deposited as the result of active transcription, thereby explaining the uni- or bimodal patterning observed in the plants and animals

studied here, respectively. This has been shown most recently by Wang et al. [75], in a paper demonstrating that deposition of H3K4me3 and H3K27ac are both dependent upon continued transcription by Pol II. These results are consistent with the idea that PTMs are so strongly correlated with transcriptional activity because they are the effect, not the cause, and instead provide a supportive role for continual transcription, rather than playing a role in its primary initiation.

However, it is clear that the chromatin landscape and transcriptional directionality are intrinsically linked. It is worth considering whether the presence or absence of a particular PTM or variant may be necessary in order to promote the initiation of transcription. This could suggest that the unilateral distribution of a PTM or variant on one side of the TSS prevents bidirectional transcription from occurring in *Arabidopsis*. The histone demethylase Flowering Locus D has been found to help limit transcription in regions of convergent genes to prevent accumulation of antisense RNAs by downregulating the presence of H3K4me1 [76]. Similarly, yeast studies have shown that the Hda1 histone deacetylase complex represses divergent transcription by deacetylating histone H3 [77]. This builds evidence towards the idea that plants precisely regulate both transcription of overlapping RNAs and likely, bidirectional transcription because antisense RNAs could trigger RNAi, and cause improper gene silencing.

Finally, it is also important to consider the presence and positioning of histone variants, such as H2A.Z, and its potential role in transcription regulation and initiation. In animals, H2A.Z is found at well-positioned -1 and +1 nucleosomes at actively transcribing genes, while in Arabidopsis, it is generally only found at the +1 nucleosome. It has been proposed that presence of H2A.Z lowers the barrier for Pol II initiation thereby allowing for transcription, with the caveat that PTMs such as acetylation or monoubiquitination add further nuance, either promoting or

repressing transcription, respectively. While these observations alone do not reveal whether or not H2A.Z presence is a pre-requisite or a result of transcription, future studies such as this one that utilize cross-species comparison to explore the connection between H2A.Z deposition, position, modification, and the necessity or sufficiency of its presence to promote initiation, will be highly informative.

Supplementary Figures
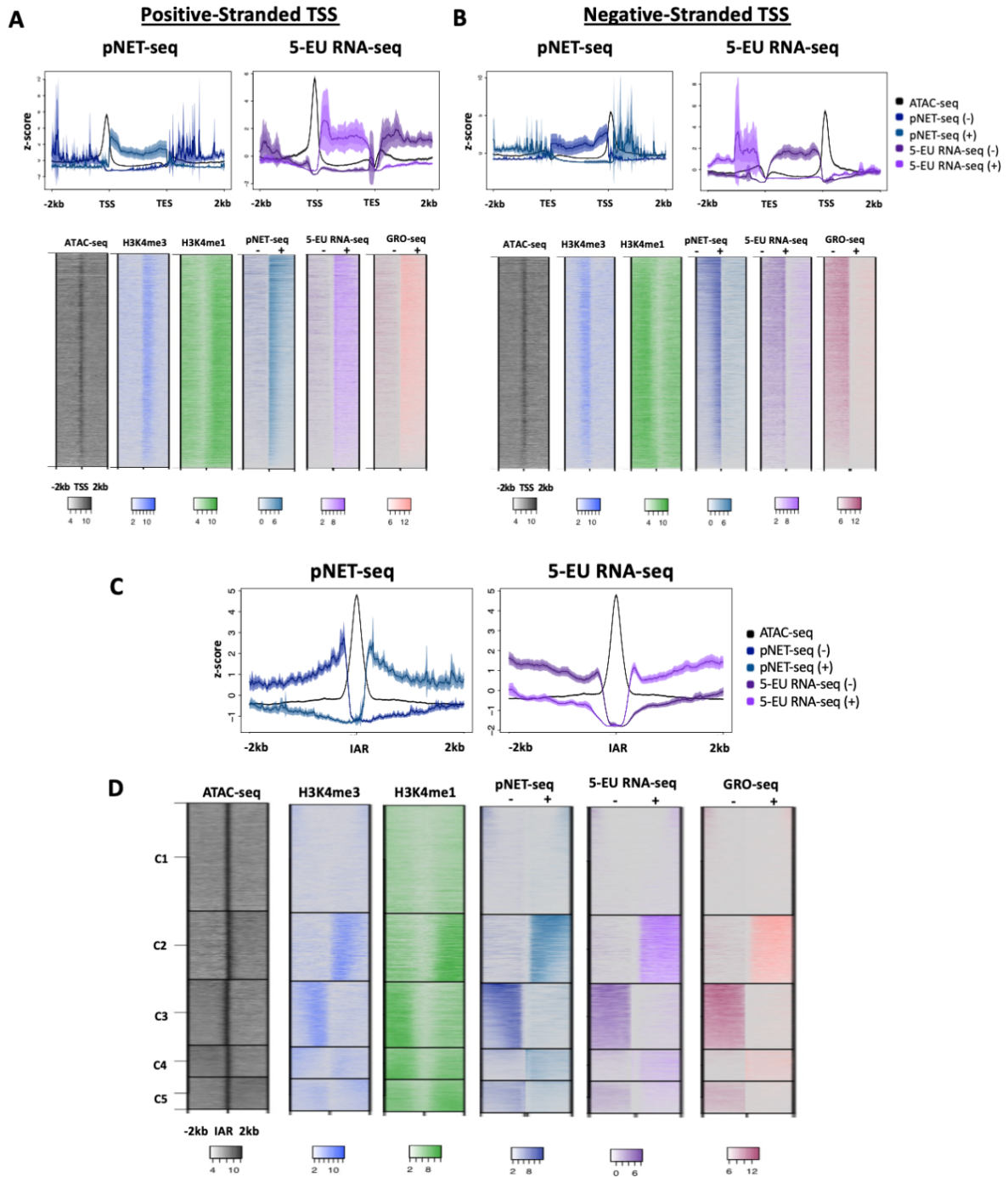
Supplementary Figure 1: Additional nascent RNA-seq datasets at TSSs and IARs in Arabidopsis thaliana. 5-EU RNA-seq and pNET-seq are shown as average plots over gene bodies (top images) at Arabidopsis positive-stranded TSSs (A) as well as negative-stranded TSSs (B).

Solid lines represent the mean signal intensity (z-score transformed); the inner, dark-shaded region represents the standard error; 6 and the outer, light-shaded region represents the 95% confidence interval (CI) of signal intensity. Lower images in each panel show the nascent RNA data in heatmap form, along with ChIP-seq data for H3K4me1 and H3K4me3. (C). Average plots of nascent RNA signals at Arabidopsis intergenic accessible chromatin regions (accessible sites outside of gene bodies and 100-2000 bp upstream of a TSS). (D) Images in each panel are K-means clustered heatmaps containing all three types of nascent RNA data, along with ChIPseq data for H3K4me1 and H3K4me3 at intergenic accessible regions.
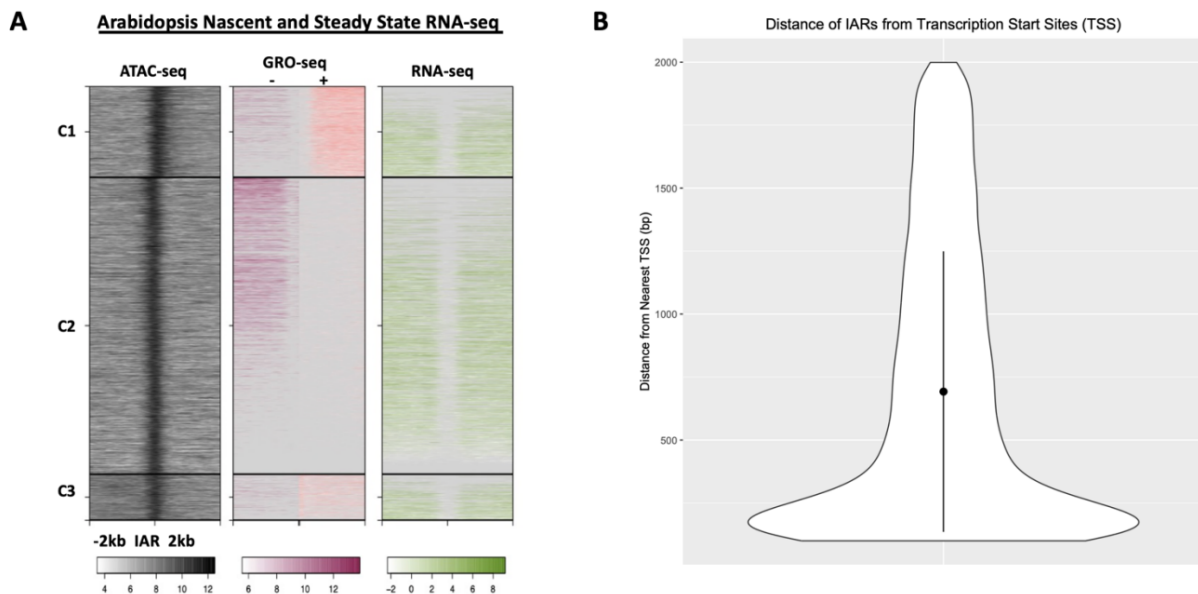
Supplementary Figure 2: Comparison of nascent and steady-state RNA-seq at intergenic enhancer regions in Arabidopsis. A) Clustered heatmaps of ATAC-seq, GRO-seq, and RNA-seq at intergenic accessible regions (IARs), which are defined as accessible chromatin sites outside of transcribed protein coding genes and within the range of 100-2000 bp from the nearest TSS. Windows extend 2 kb upstream and 2 kb downstream of the IARs; heatmaps have been log2

transformed. B) Distribution of distances between IARs and TSSs, showing an average distance

of 692 bp from a TSS.

Supplementary Tables

Supplementary Table 1: Publicly available dataset information

Species	Sample Type	Data Type	Data Source	Experiment Accession Number	Accession Number(s) of Raw Files Downloaded	File Type	Genome Version Used
Arabidopsis thaliana	Root epidermal non-hair cell nuclei	ATAC-seq	GEO	GSE101482	GSM2704265	.fastq	TAIR10
Arabidopsis thaliana	6-day old seedlings	GRO-seq	GEO	GSE83108	GSM2193124; GSM2193125	.fastq	TAIR10
Arabidopsis thaliana	12-day-old seedlings	pNET-seq	GEO	GSE109974	GSM2974949; GSM2974950	.fastq	TAIR10
Arabidopsis thaliana	5-day-old seedlings	5-EU RNA-seq	GEO	GSE118462	GSM3330492	.fastq	TAIR10
Arabidopsis thaliana	3rd and 4th rosette leaves	RNA-seq	*To be deposited (available on request)*				TAIR10
Homo sapiens	Common myeloid progenitor cell, CD34-positive female adult (27 yrs.)	ChIP-seq, H3K27Ac	ENCODE	ENCSR891KSP	ENCFF668OEU ; ENCFF641LUF ; ENCFF904CSC	.fastq	GRCh38.90
Homo sapiens	Common myeloid progenitor cell, CD34-positive female adult (27 yrs.)	ChIP-seq, H3K27Me3	ENCODE	ENCSR862NIZ	ENCFF279SSJ ; ENCFF399TRZ ; ENCFF410VSH ; ENCFF376VMQ ; ENCFF962LHH	.fastq	GRCh38.90
Homo sapiens	Common myeloid progenitor cell, CD34-positive female adult (27 yrs.)	ChIP-seq, H3K4Me1	ENCODE	ENCSR979YDQ	ENCFF186HNE ; ENCFF828SZM ; ENCFF886UDA ; ENCFF738ARX ; ENCFF376JZL	.fastq	GRCh38.90
Homo sapiens	Common myeloid progenitor cell, CD34-positive female adult (27 yrs.)	ChIP-seq, H3K4Me3	ENCODE	ENCSR850RTJ	ENCFF102IJI	.fastq	GRCh38.90
Homo sapiens	Common myeloid progenitor cell, CD34-positive female adult (27 yrs.)	ChIP-seq, Control for H3K4Me3	ENCODE	ENCSR707TMM	ENCFF599JOR ; ENCFF088FNX	.fastq	GRCh38.90
Homo sapiens	Common myeloid progenitor cell, CD34-positive female adult (27 yrs.)	ChIP-seq, Control for H3K4Me1, H3K27Ac, H3K27Me3	ENCODE	ENCSR919RJD	ENCFF606EYK ; ENCFF825IBW ; ENCFF054LZZ ; ENCFF168STH ;	.fastq	GRCh38.90
Homo sapiens	Common myeloid progenitor cell, CD34-positive female adult (27 yrs.)	DNase-seq	ENCODE	ENCSR122VUW	ENCFF164DKI ; ENCFF613FMP ; ENCFF776EIK ; ENCFF395CSF ; ENCFF175GQQ	.fastq	GRCh38.90
Homo sapiens	CD34+ erythrocytes	GRO-seq	GEO	GSE102819	GSM2746831 ; GSM2746829	.fastq	GRCh38.90
Drosophila melanogaster	S2 cells	DNase-seq	ENCODE	ENCSR834VXA	ENCFF005BHD	.fastq	Dm6
Drosophila melanogaster	S2 cells	ChIP-seq, H3K27Ac	GEO	GSE41440	GSM1017404 ; GSM1017405	.sra	Dm6

Drosophila melanogaster	S2 cells	ChIP-seq, H3K27Me3	GEO	GSE41440	GSM1017406	.sra	Dm6
Drosophila melanogaster	S2 cells	ChIP-seq, H3K4Me1	GEO	GSE41440	GSM1017407 ; GSM1017408	.sra	Dm6
Drosophila melanogaster	S2 cells	ChIP-seq, H3K4Me3	GEO	GSE41440	GSM1017409 ; GSM1017410	.sra	Dm6
Drosophila melanogaster	S2 cells	ChIP-seq, Control for H3K27Ac	GEO	GSE41440	GSM1017394 ; GSM1017395 ;	.sra	Dm6
Drosophila melanogaster	S2 cells	ChIP-seq, Control H3K27Me3	GEO	GSE41440	GSM1017397 ;	.sra	Dm6
Drosophila melanogaster	S2 cells	ChIP-seq, Control for H3K4Me1	GEO	GSE41440	GSM1017394 ; GSM1017397 ;	.sra	Dm6
Drosophila melanogaster	S2 cells	ChIP-seq, Control for H3K4Me3	GEO	GSE41440	GSM1017398 ; GSM1017399	.sra	Dm6
Drosophila melanogaster	S2 cells	GRO-seq	GEO	GSE23543	GSM577244	.fastq	Dm6
Oryza sativa	7-day-old leaf tissue	ATAC-seq; ChIP-seq	GEO	GSE128434	GSM3674604 ; GSM3674684 ; GSM3674685	.fastq	IRGSP-1.0
Glycine max	10-day-old leaf tissue	ATAC-seq; ChIP-seq	GEO	GSE128434	GSM3674586 ; GSM3674644 ; GSM3674645	.fastq	Glycine Max V1.0

Supplementary Table 2: *A. thaliana* ChIP-seq antibody information

Target	Antibody Name	Supplier	Concentration	Quantity Used per Reaction
H3K4me1	ab8895	Abcam	0.5 mg/mL	2 μg
H3K4me3	ab8580	Abcam	0.45 mg/mL	1.8 μg
H3K27ac	ab4729	Abcam	0.5 mg/mL	2 μg
H3K27me3	07-449	Millipore	0.5 mg/mL	2 μg
H3	ab1791	Abcam	0.5 mg/mL	2 μg

Supplementary Table 3: Data quality of *A. thaliana* root epidermal non-hair cell ChIP-seq datasets

Dataset type	Read size (nt)	Single end (SE) or paired end (PE)	Total reads (x 10^6)	Total mapped reads (x 10^6)	Total mapped q2 filtered reads (x 10^6)	Total nuclear peaks called (via HOMER)	Avg. size of peaks (bp)	Std. dev. of peak size (+/- bp)	Median size of peaks (bp)
ChIP-seq (H3K4me1)	50	SE	83.2	73.1	54.7	31,016	402.54	201.26	330
ChIP-seq (H3K4me3)	50	SE	101.1	91.5	82.6	14,718	277.12	134.57	189
ChIP-seq (H3K27ac)	50	SE	131.6	115	92	36,146	299.09	161.41	235

ChIP-seq (H3K27me3)	50	SE	22.5	19.8	16.1	27,784	378.9	223.91	303
ChIP-seq (H2A.Z)	50	SE	123.3	104.6	81.8	30,594	255.62	115.61	183
ChIP-seq (H3)	50	SE	62	53.8	34.3	23,379	254.38	76.43	211

Arabidopsis: Downstream				Arabidopsis: Upstream			
Cluster ID	Mean	Median	Standard Dev	Cluster ID	Mean	Median	Standard Dev
1	129	5.44	328	1	1661	1259	1272
2	65.6	5.44	267	2	84.4	8.74	311
3	416	151	678	3	74.9	8.74	241
4	1257	859	1133	4	145	8.74	387
5	63.2	5.44	208	5	543	204	785
Drosophila: Downstream				Drosophila: Upstream			
Cluster ID	Mean	Median	Standard Dev	Cluster ID	Mean	Median	Standard Dev
1	173	19.6	572	1	2143	1451	1983
2	131	0	665	2	59.5	0	397
3	1721	1126	1769	3	120	19.1	359
4	173	30	404	4	1277	703	1467
5	1870	1467	1348	5	1844	1475	1441
Human: Downstream				Human: Upstream			
Cluster ID	Mean	Median	Standard Dev	Cluster ID	Mean	Median	Standard Dev
1	40.1	36.6	32.5	1	9.7	0	20
2	4.44	0	12.4	2	8.24	0	18.4
3	13.4	4.65	22	3	139	119	71.1
4	69.3	61.5	35.1	4	12.1	5.62	19.2
5	183	167	65.5	5	19.9	11.2	28.6

Supplementary Table 4. Summary statistics for raw GRO-seq signal enrichment up and downstream of IARs for each species. Resulting un-transformed values of GRO-seq signal enrichment for each cluster, corresponding to Figure 2.

Materials And Methods

Publicly available datasets

Publicly accessible datasets from the Encyclopedia of DNA Elements (ENCODE) project (2004, 2012) (https://www.encodeproject.org/) and the Gene Expression Omnibus (GEO) (Edgar, Domrachev et al. 2002) (https://www.ncbi.nlm.nih.gov/geo/) were used in this study. Details about each of these datasets, including the species and cell type/tissue used, the accession numbers for each library, and the genome version that these data were mapped to in our study, are detailed in **Supplementary Table 1**.

Preparation of Arabidopsis ChIP-seq libraries

Non-hair cell nuclei were isolated from *A. thaliana* (Col-0) seedlings using INTACT as described previously (Wang and Deal 2015). ChIP-seq libraries were prepared and sequenced as described in (Adli and Bernstein 2011). The antibodies used to prepare the ChIP-seq libraries are listed in **Supplementary Table 2**.

Data analysis

Raw sequence read processing, mapping, peak calling, and genomic distribution determination were all conducted as described previously (Maher et al. 2018). Briefly, all ChIP, ATAC/DNase, and nascent RNA sequencing data was aligned to their respective reference genome using Bowtie2 (Langmead and Salzberg 2012). Reference genomes used are detailed in Supplementary Table 1. Files were then converted to .bam format and filtered for reads with a mapping quality score of 2 or higher using Samtools (Li et al. 2009). Datasets were then RPKM normalized using the bamcoverage function from deepTools (Ramírez et al. 2014) and further converted to BigWig

format for average plot and heat map visualization. Accessible region.bed files were generated by using the Bedtools (Quinlan and Hall 2010) intersect and subtract functions, either to look at the accessible regions found across all genes, or only at proximal intergenic regions 100–2000 base pairs up and downstream from known annotated genes in each species. Supplementary Table 3 details the data quality of the A. thaliana nonhair cell ChIP-seq generated by this study. Finally, for all nascent RNA-sequencing data (GRO-seq, pNET-seq, 5-EU RNA-seq), data were processed, mapped, and run through an R script which removed the top and bottom 10% of reads to prevent high-signal artifacts and noise from skewing the distribution of the metaplots. All heatmaps and metaplots were generated using SeqPlots (https://przemol.github.io/seqplots/) (Stempor and Ahringer 2016). For complete details of processing and the scripts used, please see Github.com/cgwillett/Silver-Willett-et-al-.

Calculation of Skew

Upstream and downstream bedfiles were generated by calculating the midpoint of the intergenic accessible sites from the original bedfile and then adding/subtracting 2 kb. Bigwig signal from all ChIP and GRO-seq datasets were obtained using bigWigAverageOverBed (https://www.encodeproject.org/software/bigwigaverageoverbed/) from kentUtils [78]. Results were filtered to ignore data points with no signal on either side of the IAR. Statistical significance was determined using a Wilcoxon Rank Sum Test to determine whether there were significant differences between (log2+1) transformed signal intensity for upstream and downstream regions, with a Bonferroni-corrected threshold applied to each of these [70]. Finally, data was visualized via violin plot (ggplot2) [79].

Data availability

ChIP-seq data from Arabidopsis root non-hair cells are deposited in the NCBI GEO database under accession number GSE152243. All other datasets are previously published and their accession numbers can be found in Supplementary Table 1.

1.   Strahl, B.D. and C.D. Allis, *The language of covalent histone modifications.* Nature, 2000. **403**(6765): p. 41 5.

2.   Sims, R.J., 3rd, K. Nishioka, and D. Reinberg, *Histone lysine methylation: a signature for chromatin function.* Trends Genet, 2003.   (11): p. 629 39.

3.   Barski, A., et al., *High-resolution profiling of histone methylations in the human genome.* Cell, 2007. **12** (4): p. 823 37.

4.   Liu, C.L., et al.,   -
        PLoS Biol, 2005. **3**(10): p. e328.

5.   Pokholok, D.K., et al., *Genome-wide map of nucleosome acetylation and methylation in yeast.* Cel ,   005. **122**(4): p. 517-27.

6.   *Determinants of Histone H3K4 Methylation Patterns.* Mol Cell, 2017. **68**(4): p. 773-785.e6.

7.   aterfall, and J.T. Lis,
        pausing and divergent initiation at human promoters. Science, 2008. **322**(5909): p. 1845-8.

8.   *Native elongating transcript sequencing reveals human*
                                Cell, 2015. **161**   -

9.   Szabo, E.X., et al., *Metabolic Labeling of RNAs Uncovers Hidden Features and Dynamics of the Arabidopsis Transcriptome.* Plant Cell, 2020. **32**   -

10.   Heintzman, N.D., et al., *Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome.* Nat Genet, 2007.   (3): p. 311-

11.   Neil, H., et al., *Widespread bidirectional promoters are the major source of cryptic transcripts in yeast.* Nature, 2009. **457**(7232): p. 1038-42.

   .   Seila, A.C., et al. *Divergent transcription: a new feature of active promoters.* Cell Cycl , 2009. **8**(16): p. 2557-64.

13.   Xu, Z., et al., *Bidirectional promoters generate pervasive transcription in yeast.* Nature, 2009. **457**(7232): p. 1033-7.

14.   Ro   ier F., et a   *Integrative epigenomic mapping defines four main chromatin states in Arabidopsis.* Embo j, 2011. **30**(10): p. 1928-38.

15.   Schwaiger, M., et al., *Evolutionary conservation of the eumetazoan gene regulatory landscape.* Genome Res, 2014. **24**(4):   -50.

16.   Villar, D., et al., *Enhancer evolution across 20 mammalian species.* Cell, 2015. **160**(3): p. 554 66.

17.   Zhu, B., et al., *Ge   -Wide Prediction and Validation of Intergenic Enhancers in Arabidopsis Using Open Chromatin Signatures.*   **27**   -
        2426.

18.   Weber, B., et al., *Plant Enhancers: A Call for Discovery.* Trends Plant Sci, 2016.   (11): p. 974-987.

19.   Xu, H. and T.R. Hoover, *Transcriptional regulation at a distance in bacteria.* Curr Opin Microbiol, 2001. **4**(2): p. 138-44.

20.   Berg, P.E., Z. Popovic, and W.F. Anderson, *Promoter dependence of enhancer activity.* Mol Cell Biol, 1984. **4**(8): p. 1664-8.

21.   Lee, T.I. and R.A. Young, *Transcription of eukaryotic protein-coding genes.* Annu Rev Genet, 2000. **34**: p. 77-137.
22.   Spitz, F. and E.E. Furlong, *Transcription factors: from enhancer binding to developmental control.* Nat Rev Genet, 2012. **13**(9): p. 613-26.
23.   Vernimmen, D. and W.A. Bickmore, *The Hierarchy of Transcriptional Activation: From Enhancer to Promoter.* Trends Genet, 2015. **31**(12): p. 696-708.
24.   Boyle, A.P., et al., *High-resolution mapping and characterization of open chromatin across the genome.* Cell, 2008. **132**(2): p. 311-22.
25.   Song, L. and G.E. Crawford, *DNase-seq: a high-resolution technique for mapping active gene regulatory elements across the genome from mammalian cells.* Cold Spring Harb Protoc, 2010. **2010**(2): p. pdb.prot5384.
26.   Buenrostro, J.D., et al., *Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position.* Nat Methods, 2013. **10**(12): p. 1213-8.
27.   Wang, Z., et al., *Combinatorial patterns of histone acetylations and methylations in the human genome.* Nat Genet, 2008. **40**(7): p. 897-903.
28.   Hawkins, R.D., et al., *Distinct epigenomic landscapes of pluripotent and lineage-committed human cells.* Cell Stem Cell, 2010. **6**(5): p. 479-91.
29.   Ernst, J., et al., *Mapping and analysis of chromatin state dynamics in nine human cell types.* Nature, 2011. **473**(7345): p. 43-9.
30.   Zentner, G.E., P.J. Tesar, and P.C. Scacheri, *Epigenetic signatures distinguish multiple classes of enhancers with distinct cellular functions.* Genome Res, 2011. **21**(8): p. 1273-83.
31.   Bonn, S., et al., *Tissue-specific analysis of chromatin state identifies temporal signatures of enhancer activity during embryonic development.* Nat Genet, 2012. **44**(2): p. 148-56.
32.   Pan, C.W., et al., *Functional roles of antisense enhancer RNA for promoting prostate cancer progression.* Theranostics, 2021. **11**(4): p. 1780-1794.
33.   Wang, D., et al., *Reprogramming transcription by distinct classes of enhancers functionally defined by eRNA.* Nature, 2011. **474**(7351): p. 390-4.
34.   Bernstein, B.E., et al., *The NIH Roadmap Epigenomics Mapping Consortium.* Nat Biotechnol, 2010. **28**(10): p. 1045-8.
35.   Rickels, R., et al., *Histone H3K4 monomethylation catalyzed by Trr and mammalian COMPASS-like proteins at enhancers is dispensable for development and viability.* Nat Genet, 2017. **49**(11): p. 1647-1653.
36.   Maher, K.A., et al., *Profiling of Accessible Chromatin Regions across Multiple Plant Species and Cell Types Reveals Common Gene Regulatory Principles and New Control Modules.* Plant Cell, 2018. **30**(1): p. 15-36.
37.   Zhu, J., et al., *RNA polymerase II activity revealed by GRO-seq and pNET-seq in Arabidopsis.* Nat Plants, 2018. **4**(12): p. 1112-1123.
38.   Hetzel, J., et al., *Nascent RNA sequencing reveals distinct features in plant transcription.* Proc Natl Acad Sci U S A, 2016. **113**(43): p. 12316-12321.
39.   Herz, H.M., et al., *Enhancer-associated H3K4 monomethylation by Trithorax-related, the Drosophila homolog of mammalian Mll3/Mll4.* Genes Dev, 2012. **26**(23): p. 2604-20.

40.   Core, L.J., et al., *Defining the status of RNA polymerase at promoters.* Cell Rep, 2012. **2**(4): p. 1025-35.

41.   Trinklein, N.D., et al., *An abundance of bidirectional promoters in the human genome.* Genome Res, 2004. **14**(1): p. 62-6.

42.   Kim, T.H., et al., *A high-resolution map of active promoters in the human genome.* Nature, 2005. **436**(7052): p. 876-80.

43.   Guenther, M.G., et al., *A chromatin landmark and transcription initiation at most promoters in human cells.* Cell, 2007. **130**(1): p. 77-88.

44.   Kaikkonen, M.U., et al., *Remodeling of the enhancer landscape during macrophage activation is coupled to enhancer transcription.* Mol Cell, 2013. **51**(3): p. 310-25.

45.   Kapranov, P., et al., *RNA maps reveal new RNA classes and a possible function for pervasive transcription.* Science, 2007. **316**(5830): p. 1484-8.

46.   Seila, A.C., et al., *Divergent transcription from active promoters.* Science, 2008. **322**(5909): p. 1849-51.

47.   Kim, T.-K., et al., *Widespread transcription at neuronal activity-regulated enhancers.* Nature, 2010. **465**(7295): p. 182-187.

48.   Hah, N., et al., *Enhancer transcripts mark active estrogen receptor binding sites.* Genome Res, 2013. **23**(8): p. 1210-23.

49.   Shlyueva, D., G. Stampfel, and A. Stark, *Transcriptional enhancers: from properties to genome-wide predictions.* Nat Rev Genet, 2014. **15**(4): p. 272-86.

50.   Tsompana, M. and M.J. Buck, *Chromatin accessibility: a window into the genome.* Epigenetics Chromatin, 2014. **7**(1): p. 33.

51.   Jiang, J., *The 'dark matter' in the plant genomes: non-coding and unannotated DNA sequences associated with open chromatin.* Curr Opin Plant Biol, 2015. **24**: p. 17-23.

52.   Bell, O., et al., *Determinants and dynamics of genome accessibility.* Nat Rev Genet, 2011. **12**(8): p. 554-64.

53.   Lu, Z., et al., *The prevalence, evolution and chromatin signatures of plant regulatory elements.* Nat Plants, 2019. **5**(12): p. 1250-1259.

54.   Jores, T., et al., *Identification of Plant Enhancers and Their Constituent Elements by STARR-seq in Tobacco Leaves.* Plant Cell, 2020. **32**(7): p. 2120-2131.

55.   Schones, D.E., et al., *Dynamic regulation of nucleosome positioning in the human genome.* Cell, 2008. **132**(5): p. 887-98.

56.   Henikoff, S., et al., *Genome-wide profiling of salt fractions maps physical properties of chromatin.* Genome Res, 2009. **19**(3): p. 460-9.

57.   Jin, C., et al., *H3.3/H2A.Z double variant-containing nucleosomes mark 'nucleosome-free regions' of active promoters and other regulatory regions.* Nat Genet, 2009. **41**(8): p. 941-5.

58.   Bagchi, D.N. and V.R. Iyer, *The Determinants of Directionality in Transcriptional Initiation.* Trends Genet, 2016. **32**(6): p. 322-333.

59.   Zhang, Y., et al., *Dynamic enhancer transcription associates with reprogramming of immune genes during pattern triggered immunity in Arabidopsis.* BMC Biol, 2022. **20**(1): p. 165.

60.   Shilatifard, A., *Chromatin modifications by methylation and ubiquitination: implications in the regulation of gene expression.* Annu Rev Biochem, 2006. **75**: p. 243-69.

61. Li, B., M. Carey, and J.L. Workman, *The role of chromatin during transcription.* Cell, 2007. **128**(4): p. 707-19.

62. Creyghton, M.P., et al., *Histone H3K27ac separates active from poised enhancers and predicts developmental state.* Proc Natl Acad Sci U S A, 2010. **107**(50): p. 21931-6.

63. Rada-Iglesias, A., et al., *A unique chromatin signature uncovers early developmental enhancers in humans.* Nature, 2011. **470**(7333): p. 279-83.

64. Bae, S. and B.J. Lesch, *H3K4me1 Distribution Predicts Transcription State and Poising at Promoters.* Front Cell Dev Biol, 2020. **8**: p. 289.

65. Melgar, M.F., F.S. Collins, and P. Sethupathy, *Discovery of active enhancers through bidirectional expression of short transcripts.* Genome Biol, 2011. **12**(11): p. R113.

66. Larke, M.S.C., et al., *Enhancers predominantly regulate gene expression during differentiation via transcription initiation.* Mol Cell, 2021. **81**(5): p. 983-997 e7.

67. Murphy, Z.C., et al., *Regulation of RNA polymerase II activity is essential for terminal erythroid maturation.* Blood, 2021. **138**(18): p. 1740-1756.

68. Andersson, R., et al., *An atlas of active enhancers across human cell types and tissues.* Nature, 2014. **507**(7493): p. 455-461.

69. Lin, T., et al., *Extending the Mann-Whitney-Wilcoxon rank sum test to survey data for comparing mean ranks.* Stat Med, 2021. **40**(7): p. 1705-1717.

70. Curtin, F. and P. Schulz, *Multiple correlations and Bonferroni's correction.* Biol Psychiatry, 1998. **44**(8): p. 775-7.

71. Liu, C., et al., *Genome-wide analysis of chromatin packing in Arabidopsis thaliana at single-gene resolution.* Genome Res, 2016. **26**(8): p. 1057-68.

72. Tan-Wong, S.M., et al., *Gene loops enhance transcriptional directionality.* Science, 2012. **338**(6107): p. 671-5.

73. Castelnuovo, M. and F. Stutz, *Gene loops and HDACs to promote transcription directionality.* Nucleus, 2013. **4**(2): p. 92-4.

74. Wu, C.-H., et al., *NELF and DSIF cause promoter proximal pausing on the hsp70 promoter in Drosophila.* Genes & development, 2003. **17**(11): p. 1402-1414.

75. Wang, Z., et al., *Prediction of histone post-translational modification patterns based on nascent transcription data.* Nature Genetics, 2022. **54**(3): p. 295-305.

76. Inagaki, S., et al., *Chromatin-based mechanisms to coordinate convergent overlapping transcription.* Nature Plants, 2021. **7**(3): p. 295-302.

77. Gowthaman, U., et al., *The Hda1 histone deacetylase limits divergent non-coding transcription and restricts transcription initiation frequency.* Embo j, 2021. **40**(23): p. e108903.

78. Kent, W.J., et al., *BigWig and BigBed: enabling browsing of large distributed datasets.* Bioinformatics, 2010. **26**(17): p. 2204-7.

79. Wickham, H., *ggplot2: Elegant Graphics for Data Analysis.* Springer-Verlag New York, 2016.

Chapter 3

Measuring genome-wide dynamics of H3.3 turnover in *Arabidopsis thaliana*

Chapter 3: Measuring genome-wide dynamics of H3.3 turnover in *Arabidopsis thaliana*

Courtney G. Willett[1,2] and Roger B. Deal[1]

[1]Department of Biology

[2]Graduate Program in Genetics and Molecular Biology

Abstract

The nucleosome is the core unit of chromatin organization in eukaryotes. Comprised of an octamer of four core histones (H2A, H2B, H3, H4), the nucleosome plays a key regulatory role in DNA-templated processes, such as DNA replication and transcription. Sequence variants of these core histone proteins can be incorporated into the genome in a replication-independent manner and influence the surrounding epigenetic and transcriptional states. Of note, histone variant H3.3 is specifically associated with euchromatic regions of the genome. These regions include genes actively undergoing transcription, which disrupts the nucleosome. The disassembly and reassembly of the nucleosome in wake of transcription, or nucleosome turnover, results in histones being ejected, recycled and/or replaced by newly synthesized histones. However, a system to measure the dynamics of this in real time has yet to be created in plants. We have developed and characterized a transgenic Arabidopsis thaliana line carrying an estradiol-inducible GFP-tagged H3.3. Using this system for H3.3 induction followed by ChIP-seq, we observed a positive correlation between gene transcriptional activity and H3.3-GFP incorporation, with new H3.3 being preferentially incorporated at the 3' end of active genes. This is consistent with what has been observed in steady-state H3.3 profiling experiments in plants, but futher reveals the dynamics of H3.3 incorporation. Overall, this system has the potential to measure H3.3 dynamics in a variety of environmental conditions and mutant backgrounds,

allowing us to address questions regarding the maintenance or erasure of epigenetic marks during transcription.

Introduction

The eukaryotic genome is organized according to the morphology and function for each cell type within a multicellular organism. It is tightly compacted within the nucleus into repeating nucleosome units, each consisting of ~147 bp of DNA wrapped around an octameric histone protein core [1]. The histone core contains a tetramer of H3-H4 proteins, flanked by two H2A-H2B dimers. Each histone can be modified by distinct post-translational modifications (PTMs) [2]. The interactions between histone proteins themselves, PTMs, and surrounding DNA contribute to changes in the chromatin landscape.

Notably, the introduction of histone variants into the nucleosome can change its stability and alter chromatin accessibility. While the canonical histones are primarily incorporated during DNA replication, histone variants can be incorporated outside of S phase and are considered replication-independent [3]. These histone variants influence the internal stability of the nucleosome, thus altering the surrounding chromatin landscape and providing key insight into the relationship between eukaryotic epigenetics and chromatin dynamics [3].

During S phase of the cell cycle, the process of DNA replication doubles the demand for nucleosomes, and new histones are synthesized to reassemble chromatin during replication. The disassembly and reassembly of nucleosomes, or turnover, is mediated by histone chaperones, a diverse group of proteins that aid in incorporation, eviction and modification of histones [4, 5]. Outside of S phase, nucleosome turnover results in histone variants being incorporated at specific

genomic regions [5]. Specifically, histone variant H3.3 undergoes high rates of incorporation at transcriptionally active genes, proportional to their expression level [6-11] .

Histone H3 replacement dynamics are generally similar between animals and plants. In both, histone H3 is encoded by three variants: replicative H3.1/H3.2, replication-independent H3.3, and centromeric H3 (CenH3) [12]. H3.1 and H3.3 differ only at four amino acids in flowering plants but have distinct deposition patterns and behaviors. During replication, H3.1 is evenly distributed throughout the genome [5, 13, 14]. However, outside of S phase, nucleosomes can be disrupted during transcription and H3.3 is then able to be incorporated into the nucleosome, as H3.1 is not available. H3.3 is associated with euchromatic transcribed chromatin in both animals and plants [10, 15, 16]. H3.3 is enriched mainly at the 3' end of actively transcribed genes in Arabidopsis, overlapping with enrichment of RNA Polymerase II [16, 17]. Historically, it was thought that H3.3 cannot be methylated at K27 by the Polycomb Repressive Complex (PRC) due to its association with transcriptionally active genes [3, 10, 16]. However, recent data suggests otherwise; moreover, H3K27me3 on H3.3 plays a key regulatory role in developmental gene expression in Arabidopsis [18]. This underscores our growing understanding of the role of H3.3 in regulating gene expression and the need to further investigate its role in maintaining euchromatic regions of the genome.

However, while total loss of H3.3 is lethal, partial knockdown of H3.3 in *Drosophila* [19], mouse embryonic stem cells (ESCs) [20, 21] and Arabidopsis [22] does not have the expected global effect on general transcription or H3K4me3 levels. Instead, H3.3 prevents the deposition of opposing histone variants and causes changes in DNA methylation. The loss of H3.3 at transcribed genes results in reduced DNA methylation, as well as H2A.Z and H1 occupancy in regions previously occupied by H3.3 [22].

Turnover dynamics of H3.3 may help us better understand the nature of its relationship with transcription activation and the stability of epigenetic marks. However, a system to measure H3.3 incorporation in real-time is needed to study this. Transcription disrupts the nucleosome and results in greater rates of H3.3 incorporation[10]. Previous studies in cultured cells (mouse, Drosophila) and yeast found that regions of active gene expression are unusually dynamic and have much higher rates of nucleosome turnover and H3.3 incorporation compared to repressed regions [9, 23]. Additionally, it has been observed that H2A/H2B dimers turn over rapidly within promoter and gene bodies, regardless of expression level, indicating a distinction from the turnover of the H3/H4 tetramer [24]. However, this dynamic is less understood in plants.

While H3.3 knockdown studies in plants did not observe global transcriptional changes, there were differences in expression in environmental-response genes, suggesting that H3.3 dynamics are more specific to these response genes [22]. Environmental stress can induce activation of stress response genes, thereby altering DNA methylation, histone PTMs and chromatin structure. Upon recurrence of these stress events, the plant is able to adapt much quicker in a mechanism referred to as "transcriptional stress memory"[25]. H3.3 occupancy and H3K4me3 are associated with transcriptional stress memory, such as response to heat stress in plants [26]. In keeping with this observed relationship between H3.3 and transcription, a unique pattern is observed at heat stress-induced memory genes, wherein rates of H3.3 incorporation are much lower compared to non-memory genes [27]. Previous studies observed a correlation between histone turnover and maintenance of histone methylation, wherein differential turnover has effects on retention of specific histone modifications, depending on the mechanism by which the marks are written and erased [27-29]. H3K4me3 is a mark of recent transcriptional activity and nucleosomes modified with H3K4me3, specifically at stress-response genes, were observed to have lower rates of

turnover during a stress response [27]. Hence, genes that are highly transcribed in response to stress are able to retain H3K4me3-marked histones, as they are not being ejected and replaced during nucleosome turnover [27]. This differential epigenetic mark retention may therefore underlie the phenomenon of transcriptional stress memory.

In order to examine patterns of H3.3 turnover dynamics genome-wide in Arabidopsis, we created an estradiol-inducible GFP-tagged H3.3 in a wild-type background. Previous studies have speculated a positive correlation between H3.3 turnover and gene expression, however the dynamics of this relationship in plants have yet to be measured globally. Our results describe a system for estimating H3.3 turnover genome-wide. We observe a positive correlation between transcript levels and rate of H3.3-GFP incorporation, consistent with what has been observed in yeast and cell culture studies but with distinct dynamics in relation to the gene body. While nucleosome turnover and H3.3 incorporation is observed across transcribed gene bodies in fungi and animals, this process in plants is restricted to the distal 3' ends of genes.

Results

To study histone H3.3 turnover dynamics in real time in *Arabidopsis thaliana*, we generated a transgenic plant expressing GFP-tagged H3.3 (HTR5) under control of an estradiol-inducible promoter. H3.3 differs from canonical H3 only by four amino acids, creating a need for a tagged-H3.3, as a specific antibody for H3.3 is not available [3]. Previous papers have successfully utilized GFP-tagged H3.3 to assess steady-state H3.3 incorporation in plants [16, 30]. Thus, despite being epitope-tagged, tagged H3.3 dynamics appear to mirror those of native H3.3 [9, 15].

We first wanted to test the induction dynamics of our transgene by conducting a 24 hour time course to observe the rate of GFP expression after induction by estradiol. Expression of GFP under estradiol treatment was assessed in the roots of 7-day-old plants. Both wildtype and H3.3-GFP plants were grown on half-strength Murashige and Skoog (1/2 MS) agar plates for 7 days and then induced with estradiol for 24 hours. Plants were then moved back to MS plates without estradiol for an additional 24 hours to monitor disappearance of the GFP signal. We observed GFP signal first at 5 hours of induction within the H3.3-GFP plants, and this signal grew stronger over the course of 24 hours (Fig 1A). We next wanted to test whether GFP expression would decrease following removal of estradiol. However, removal of the estradiol was not sufficient to reduce GFP signal in the transgenic plants even after 24 hours (Fig 1B). This inducible system could hypothetically be used to either measure the H3.3-GFP incorporation rate or the loss of H3.3-GFP (i.e. following the removal of estradiol); however, the latter is only possible if the GFP protein is degraded after removal of estradiol. Ergo, moving forward we decided to measure the rate at which H3.3-GFP is incorporated after induction.

We next tested incorporation of H3.3-GFP into chromatin by probing chromatin-associated H3.3-GFP via Western Blotting, as previously described [31]. Chromatin fractions were isolated from H3.3-GFP transgenic plants that were harvested pre- and post-estradiol induction (24 hours). Anti-GFP antibodies were able to recognize H3.3-GFP in the induced sample, suggesting that H3.3-GFP was successfully incorporated into chromatin (Fig 1C). Moreover, anti-H3 is not able to detect H3.3-GFP, most likely because anti-H3 recognizes the N-terminus of H3 which is where GFP is fused to our transgenic H3.3 (Fig 1C).
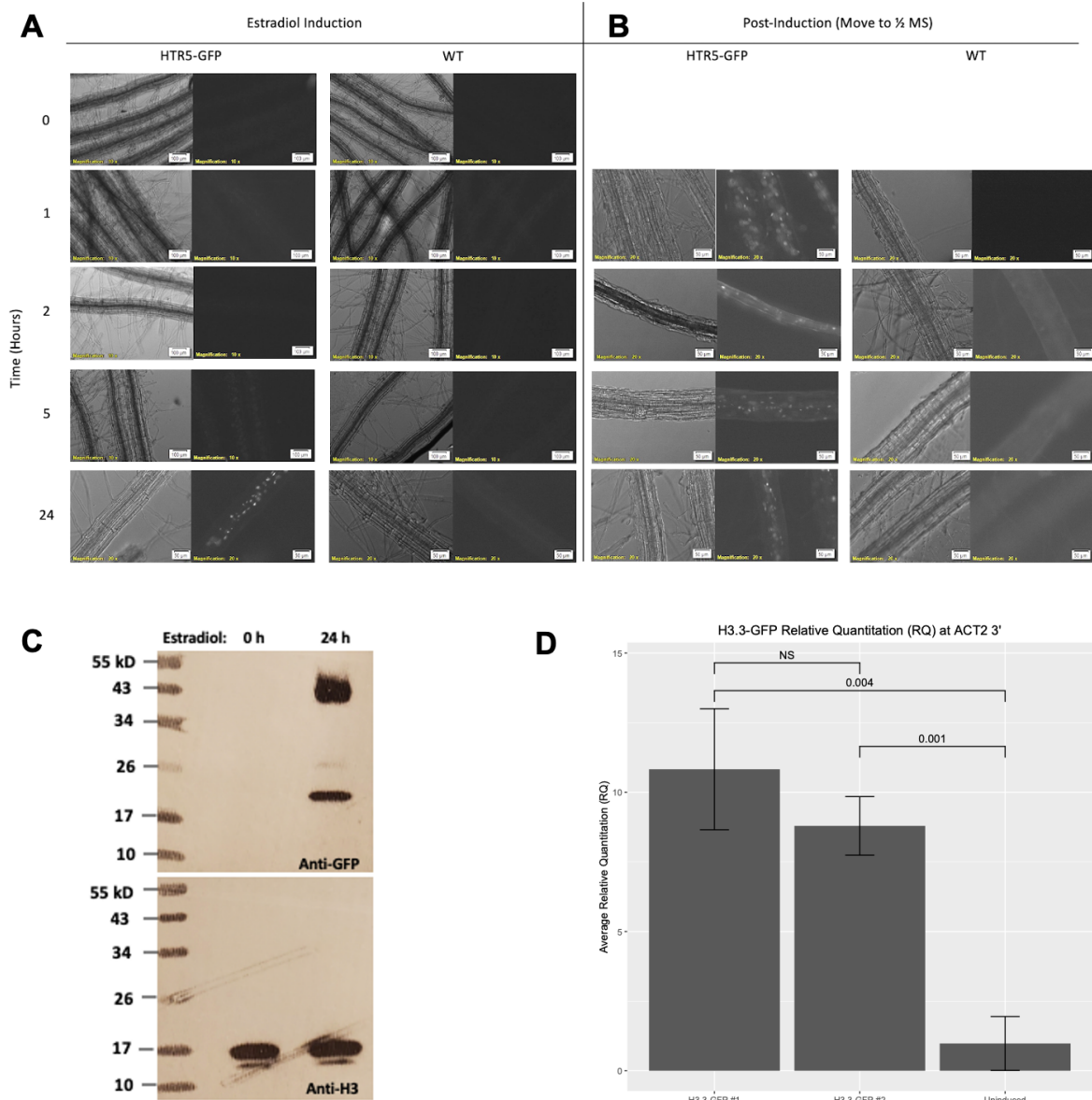
Figure 1: H3.3-GFP is detectable after 5 hours of induction and strongest at 24 hours. A) Fluorescence microscopy reveals GFP fluorescence is observable after 5 hours of estradiol induction. B) Removal of estradiol is not sufficient to reduce GFP signal, even after 24 hours of no induction. C) Western blots of insoluble chromatin fractions from 0 and 24 hour-induced samples revealed that H3.3-GFP (~42 kD) is successfully incorporated into chromatin. D) ChIP-qPCR of H3.3-GFP signal is significantly (p<0.05) increased at the 3' end of ACT2 after 24 hours of induction, suggesting active incorporation of H3.3-GFP at ACT2.

Finally, before genome-wide profiling, we conducted ChIP-qPCR on the 3' end of the ACTIN 2 (ACT2) gene, using two different estradiol-inducible H3.3-GFP transgenic lines. H3.3 is enriched across the gene bodies of actively transcribed genes, with a bias towards the 3' end [3, 10, 16]. Thus, we expected to see dynamic deposition of H3.3-GFP at the 3' end of ACT2, an actively transcribed gene, after 24 hours of estradiol induction. Indeed, H3.3-GFP is not detectable pre-induction, and significantly ($p<0.05$) increases after 24 hours of induction (Fig 1D).

We next tested H3.3-GFP incorporation genome-wide using enhanced ChIP-seq [32]. Plants were grown vertically on MS plates for 10 days and then induced with estradiol for 24 hours; root samples were then taken at 5 and 24 hours of induction for ChIP-seq using a GFP antibody. Metaplots of our inducible H3.3-GFP ChIP-seq signal at 5 and 24 hours of estradiol induction, as well as publicly available constitutively expressed H3.3-GFP [16], revealed the expected pattern of H3.3 incorporation over gene bodies. First, the steady-state H3.3-GFP data showed H3.3 enrichment over the gene bodies, with a preference for the 3' end (Fig 2A, blue). The inducible H3.3 showed enrichment at the 3' end only (Fig 2A, red and green); however this is to be expected as we are measuring dynamically deposited H3.3, which occurs at the 3' end whereas gene body H3.3 is more stable [3, 16]. However, as the transgene is induced between 5, 10 and 24 hours, we observe some encroaching of the H3.3-GFP into the gene body from the 3' end, suggesting that eventually the inducible H3.3-GFP pattern will match steady-state H3.3 patterns over the gene body (Fig 2A).
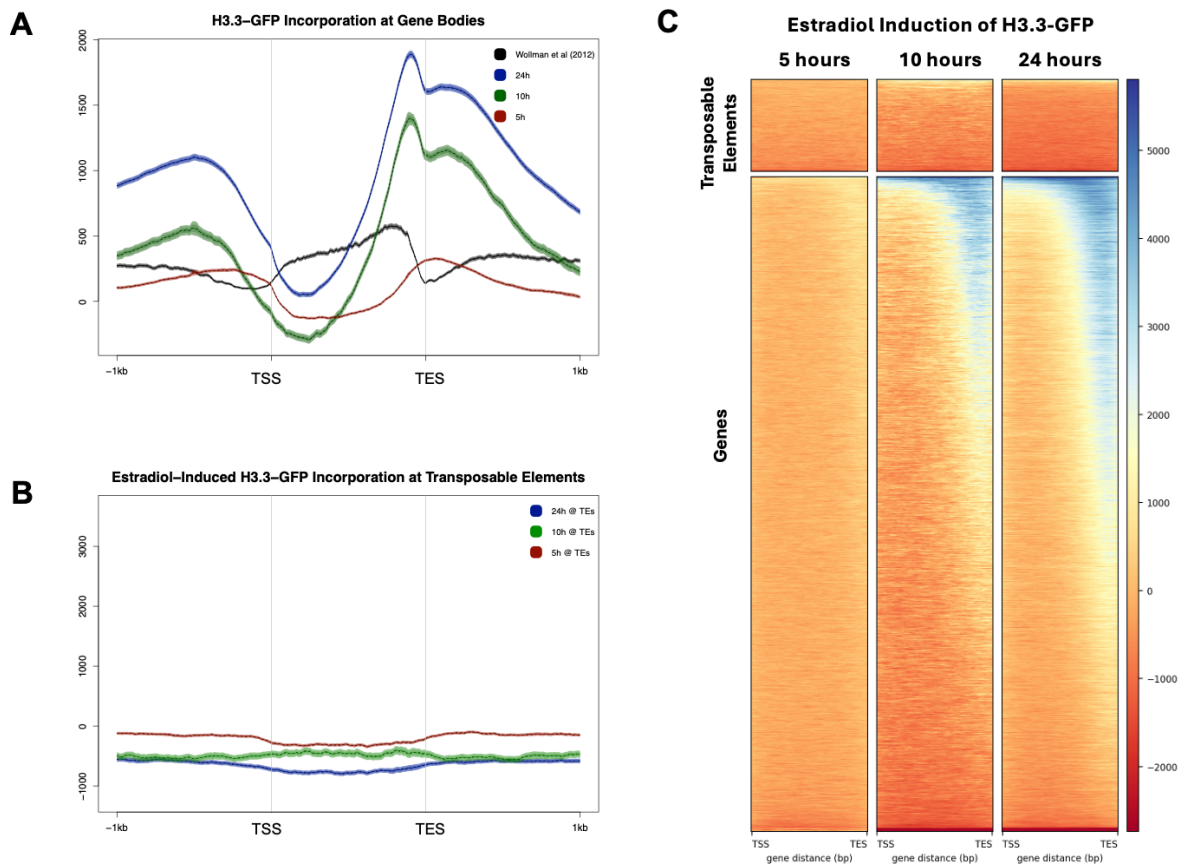
Figure 2: H3.3-GFP is turned over preferentially at the 3' end of gene bodies. A) Metaplot of H3.3-GFP after 5 (red), 10 (green), and 24 (blue) hours of estradiol induction, compared to publicly available H3.3-GFP (black), over gene bodies. There is a substantial difference between 5 and 24 hours, suggesting H3.3-GFP is being dynamically deposited at genic sites, with a preference for the 3' end. B) Compared to genes, H3.3-GFP does not appear to be incorporated at transposable elements (TEs). C) Heatmaps show no increase in H3.3-GFP between 5 and 24 hours of estradiol induction at TEs, but there is a marked increase at genes.

H3.3 shows a preference for transcribed genes [3, 10, 16]. As such, we next plotted H3.3-GFP enrichment at gene bodies and transposable elements, which are transcriptionally repressed (Fig 2B). Similar to Figure 2A, there is a substantial increase in H3.3-GFP enrichment at the 3' end of genes between 5 and 24 hours of induction; however, transposable elements (TEs) do not share this pattern (Fig 2B). In fact, it appears that H3.3-GFP enrichment decreases from 5 to 24 hours at TEs, but this is likely an artifact due to very low signal rather than a true biological phenomenon.

Heatmaps of our inducible H3.3-GFP at TEs and gene bodies recapitulate the relationship between H3.3 deposition and euchromatic regions. There is virtually no H3.3-GFP at TEs at either 5 or 24 hours of estradiol induction, whereas there is a marked increase of H3.3-GFP incorporation at genes (Fig 2C).

To further assess the relationship between transcription and H3.3 incorporation, we examined the rate of H3.3 incorporation between 5 and 24 hours of estradiol induction at genes ranked by transcript level based on RNA-seq data [33]. At 5 hours of induction, there is not an observable difference in H3.3 incorporation between differentially transcribed genes (Fig 3A). However, at 24 hours there is a clear positive correlation between transcript level and H3.3 incorporation (Fig 3B-C). Higher expressed genes have greater enrichment of H3.3 incorporation at the 3' end and seem to show a pattern of H3.3 steadily encroaching upstream into the gene body (Fig 3C). Thus, more highly expressed genes have greater rates of H3.3 incorporation and this incorporation extends further into the gene body.
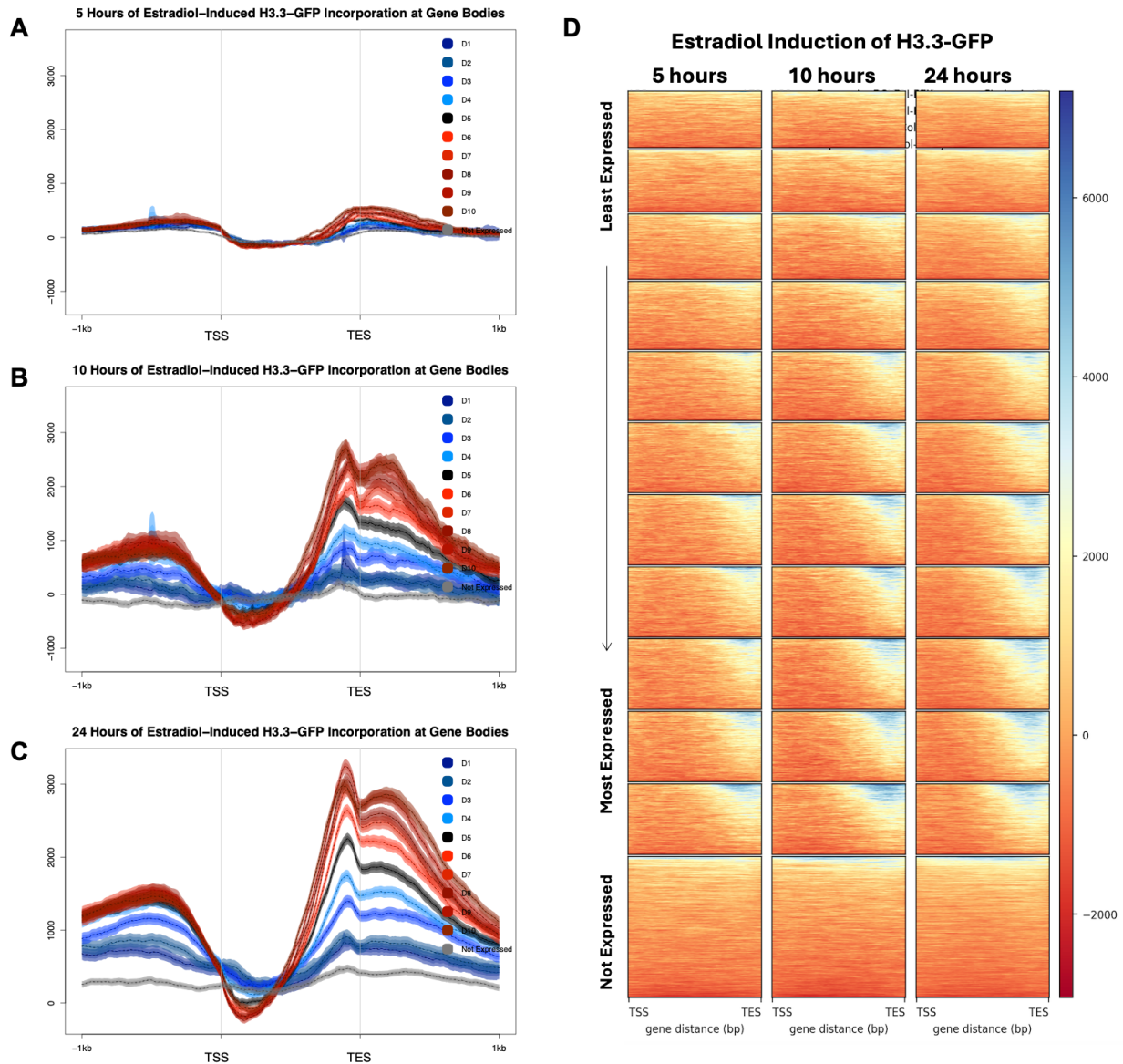
Figure 3: H3.3-GFP incorporation is positively correlated with gene expression level. A) Metaplot of H3.3-GFP signal at 5 hours of induction shows no difference between gene expression level. B) Metaplots of the same data at 24 hours of induction reveals a positive correlation between enrichment of H3.3-GFP and gene expression level, suggesting greater turnover of H3.3 is relative to gene expression level. C) Heatmaps of H3.3-GFP incorporation between 5 and 24 hours reveal more actively transcribed genes have greater incorporation of H3.3-GFP.

H3.3 has also been observed to have high turnover at enhancer regions during differentiation in mouse embryonic stem cells [23]. To investigate whether the same pattern is observed in plants, we analyzed patterns of H3.3-GFP incorporation at intergenic accessible regions (IARs), defined as nuclease hypersensitive sites that were outside of transcribed protein coding regions [34]. Using these IARs as a proxy for enhancer identification, we observe a gradual increase in H3.3-GFP incorporation at these sites, similar to what is observed in gene bodies (Fig 4A). Indeed, there is a gradual enrichment of H3.3-GFP at these sites over the course of 24 hours of induction, indicating incorporation of H3.3 at IARs (Fig 4B).
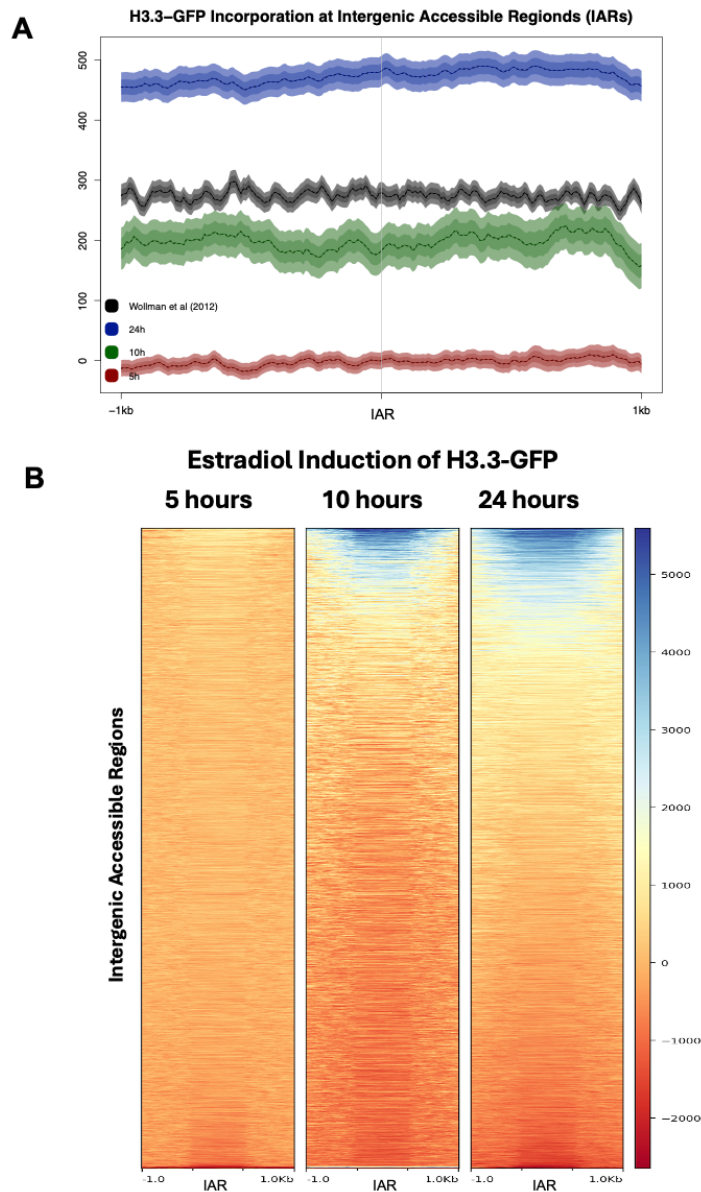
Figure 4: Estradiol-Induced H3.3-GFP incorporation at intergenic accessible regions (IARs). A) Metaplot of H3.3-GFP incorporation at 5 (red), 10 (green), and 24 (blue) hours of estradiol induction, compared to publicly available H3.3-GFP (black), at IARs. B) Heatmap of H3.3-GFP incorporation at IARs after 5, 10, and 24 hours of induction.

Previous studies have looked at H3.3 turnover rates in an environmental stress context and found that stress response genes had lower rates of H3.3 incorporation [27]. This suggests that histone retention is key to preserving epigenetic memory of these response genes and allows them to be activated quickly during a recurring stress event. Phosphate is a crucial nutrient for plant growth and as such, phosphate-starvation has been shown to illicit a strong transcriptional response in Arabidopsis [35, 36]. To simulate environmental stress, we grew plants on regular ½ MS media for 7-10 days and then transferred them to phosphate-depleted media with estradiol for 24 hours. Interestingly, we observed a delay in H3.3-GFP incorporation at 5 and 10 hours of induction on phosphate-starved media, compared to normal growth conditions (Fig 5A). While 24 hours of induction did result in H3.3-GFP incorporation at the 3' end of gene bodies, it was not as highly enriched as in normal growth conditions (Fig 5B). This is consistent with previous findings, as it appears that H3.3-GFP incorporation is greatly slowed in response to stress and therefore, H3.3 is not being turned over as quickly. Another potential explanation is that the phosphate starvation interferes with or slows the estradiol induction. However, these results are preliminary and further studies will be necessary to draw any major conclusions from this experiment on phosphate stress.
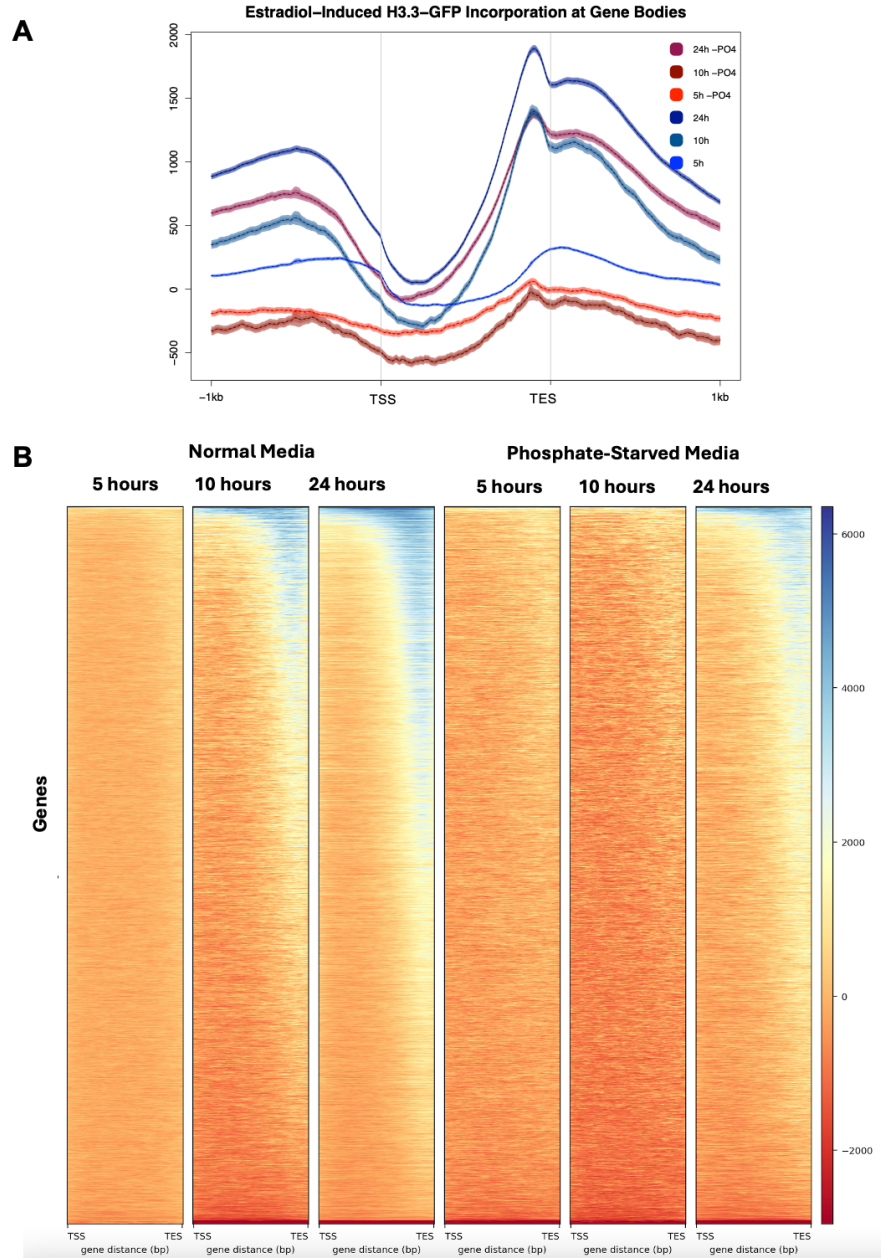
Figure 5: Phosphate-starved plants have delayed H3.3-GFP incorporation at gene bodies.

A) Metaplot of H3.3-GFP incorporation at gene bodies for 5, 10, and 24 hours of estradiol induction under normal ½ MS media (blue) or phosphate-starved media (red). All plant were grown on ½ MS media for 7-10 days; phosphate-starved plants were transferred to phosphate-deficient media with estradiol. B) Heatmap of H3.3-GFP incorporation at genes over 5, 10 and 24 hours of estradiol induction at either normal or phosphate-starved media.

Discussion

We have developed an inducible system for measuring H3.3 incorporation in plants in real-time. The relationship between histone turnover and transcription has been studied in previous models that concluded that histone turnover is higher at actively expressed genes [15, 17, 27].

The process of transcription disrupts the nucleosome, resulting in either recycling or turnover of the existing histone proteins. Nucleosome recycling is useful both for maintaining chromatin structure and epigenetic memory [5, 27]. However, more highly transcribed genes may be overwhelmed by nucleosome recycling, making nucleosome turnover a more feasible alternative [5]. In essence, nucleosome recycling may be impossible past a certain rate of transcription, as the transcriptional machinery may be overwhelmed and the chance of H3.3 incorporation increases with each round of disruption, thereby leading to greater H3.3 incorporation at these sites. We observed this phenomenon by measuring H3.3-GFP enrichment at genes, relative to their level of transcriptional activity, wherein more actively transcribed genes had greater H3.3-GFP incorporation (Fig 3C). Additionally, we observed enrichment of H3.3-GFP at intergenic accessible regions, consistent with previous findings that enhancer regions have high rates of H3.3 turnover [23].

A previous study by Baurle et al utilized a similar model to examine heat-shock inducible transcription, useful for assessing turnover in response to environmental stress [27]. Our approach is agnostic of environmental stresses and can be useful for assaying turnover under a variety of conditions. We have presented an analysis comparing nucleosome turnover in both normal and phosphate-starved conditions. Our preliminary data suggest that H3.3-GFP incorporation occurs at a much slower rate in phosphate-starved conditions compared to normal

growth conditions. This is consistent with the idea that genes responsible for environmental

stress-response undergo lower rates of H3.3 turnover to preserve epigenetic memory [27].

Moreover, previous knockdown studies of H3.3 in plant and animal models observed changes

specifically in environmental response genes, further hinting at the relationship between H3.3

and stress response gene activation [12, 22].

Overall, we have developed an inducible system for measuring H3.3 incorporation

dynamics in real-time. This has allowed us to visualize the correlation between gene

transcriptional activity in plants under normal and environmental stress conditions and will be a

highly useful tool for further investigations.

Materials and Methods

Plant Materials

A.thaliana (Col-o) plants were grown on half-strength Murashige and Skoog plates containing

0.8% agar and 1% sucrose for 7-10 days. Roots were incubated with 10 uM estradiol for 5, 10 or

24 hours.

Generation of Transgenic Lines

The HTR5 (H3.3) coding sequence was amplified by PCR, excluding the stop codon, from a

cDNA pool and was cloned into the pENTR-D-TOPO vector by topoisomerase cloning. A

sequence-verified clone of the HTR5 coding sequence, minus the stop codon, was then cloned

into the pMDC7 vector by Gateway cloning to place the HTR5 coding sequence under control of

an estradiol-inucible promoter and in-frame with the downstream GFP coding sequence. A

sequence-verified clone of this HTR5-GFP construct was then transformed into Agrobacterium

tumefaciens and used to transform wild-type (Col-0 strain) Arabidopsis plants. Multiple

independent transgenic lines were isolated and bred to homozygosity for the estradiol-inducible

HTR5-GFP transgene.

Chromatin Isolation and Western Blot

Chromatin was isolated from 7–10-day-old *A. thaliana* roots as previously described [31]. In

brief, tissue was ground in liquid nitrogen and homogenized with 5 mL Honda Buffer (20 mM

HEPES–KOH, pH 7.4, 0.44 M sucrose, 1.25% Ficoll, 2.5% Dextran T40, 10 mM MgCl2, 0.5%

Triton X-100, 5 mM DTT, 1 mM PMSF, and proteinase inhibitor cocktail, Roche). The sample

was then filtered, centrifuged, and the pellet washed with Honda buffer and 1X PBS (1 mM

EDTA) to produce an insoluble fraction containing chromatin. This pellet was resuspended in

protein loading dye for western blotting and probed with anti-H3 and anti-GFP.

Preparation of Arabidopsis ChIP-seq libraries

ChIP-seq libraries were prepared and sequenced as described previously [32]. In brief, roots of

estradiol-induced plants were crosslinked with 1% formaldehyde and 0.01% silwet L-77,

followed by quenching with 0.2 M glycine. Samples were ground in liquid nitrogen in Buffer S

(50 mM HEPES-KOH pH 7.5, 150 mM NaCl, 1 mM EDTA, 1% Triton X-100, 0.1% sodium

deoxycholate, 1% SDS, and 1X Complete protease inhibitor). They were next sonicated for 1

hour at 4 degrees (45 seconds on/15 seconds off). Sonicated lysates were spun down and the

supernatant was transferred to a tube containing 10 mL of Buffer F (50 mM HEPES-KOH pH

7.5, 1 mM EDTA, 1% Triton X-100, 0.1% sodium deoxycholate and 1X Complete protease

inhibitor). Immunoprecipitation of H3.3-GFP was done using ChomoTek GFP-Trap Magnetic

Agarose beads (AB_2631358) for 2 hours at 4 degrees. The beads were washed for 5 minutes twice with 1 mL low-salt ChIP wash buffer (150 mM NaCl), twice with high-salt ChIP wash buffer (350 mM NaCl), once with LiCl wash buffer (250 mM LiCl), and twice with TE buffer (10 mM Tris pH=8). Finally, proteins were eluted from the beads using ChIP Elution Buffer (50 mM Tris-HCL pH=7.5, 10 mM EDTA, 1% SDS) for 15 minutes at 65 degrees. Eluted proteins were incubated with proteinase K overnight at 55 degrees to reverse crosslinking. DNA was purified using the Qiagen Minelute Kit (Cat. No. 28004). Sequencing libraries were prepared using the Takara ThruPlex DNA-seq library kit (Cat. No. R400675).

ChIP-qPCR of ACT2

Samples prepared as described above were also used for ChIP-qPCR of ACT2. Cycle threshold (Ct) values were normalized using an endogenous control (ΔCt) —in this case, a transposable element. Next, input values were subtracted from the test samples (ΔΔCt). Finally, relative quantification (RQ) was calculated by taking $2^{-\Delta\Delta Ct}$.

Sequencing and data Analysis

Libraries were sequenced on an Illumina Novaseq instrument using 150 bp paired-end reads. ChIP sequencing reads were aligned to the Arabidopsis Col-PEK genome assembly [37] using Bowtie2 [38]. Files were converted to BAM format and filtered for reads with a quality score >= 2 using SAMtools [39]. Data were normalized by Reads per Kilobase Million (RPKM) and converted to bigwig format for visualization using DeepTools [40]. Heatmaps and metaplots were generated using SeqPlots [41] and DeepTools.

References

1. Luger, K., et al., *Crystal structure of the nucleosome core particle at 2.8 A resolution*. Nature, 1997. **389**(6648): p. 251-60.
2. Xiao, J., U.S. Lee, and D. Wagner, *Tug of war: adding and removing histone lysine methylation in Arabidopsis*. Curr Opin Plant Biol, 2016. **34**: p. 41-53.
3. Foroozani, M., D.H. Holder, and R.B. Deal, *Histone Variants in the Specialization of Plant Chromatin*. Annu Rev Plant Biol, 2022. **73**: p. 149-172.
4. Hammond, C.M., et al., *Histone chaperone networks shaping chromatin function*. Nat Rev Mol Cell Biol, 2017. **18**(3): p. 141-158.
5. Robert, F. and C. Jeronimo, *Transcription-coupled nucleosome assembly*. Trends Biochem Sci, 2023. **48**(11): p. 978-992.
6. Chen, P., et al., *H3.3 actively marks enhancers and primes gene transcription via opening higher-ordered chromatin*. Genes Dev, 2013. **27**(19): p. 2109-24.
7. Jin, C., et al., *H3.3/H2A.Z double variant-containing nucleosomes mark 'nucleosome-free regions' of active promoters and other regulatory regions*. Nat Genet, 2009. **41**(8): p. 941-5.
8. Kraushaar, D.C., et al., *Genome-wide incorporation dynamics reveal distinct categories of turnover for the histone variant H3.3*. Genome Biol, 2013. **14**(10): p. R121.
9. Mito, Y., J.G. Henikoff, and S. Henikoff, *Genome-scale profiling of histone H3.3 replacement patterns*. Nat Genet, 2005. **37**(10): p. 1090-7.
10. Stroud, H., et al., *Genome-wide analysis of histone H3.1 and H3.3 variants in Arabidopsis thaliana*. Proc Natl Acad Sci U S A, 2012. **109**(14): p. 5370-5.
11. Zhao, F., et al., *The histone variant H3.3 promotes the active chromatin state to repress flowering in Arabidopsis*. Plant Physiol, 2021. **186**(4): p. 2051-2063.
12. Lin, C.J., M. Conti, and M. Ramalho-Santos, *Histone variant H3.3 maintains a decondensed chromatin state essential for mouse preimplantation development*. Development, 2013. **140**(17): p. 3624-34.
13. Escobar, T.M., A. Loyola, and D. Reinberg, *Parental nucleosome segregation and the inheritance of cellular identity*. Nat Rev Genet, 2021. **22**(6): p. 379-392.
14. Stewart-Morgan, K.R., N. Petryk, and A. Groth, *Chromatin replication and epigenetic cell memory*. Nat Cell Biol, 2020. **22**(4): p. 361-371.
15. Deal, R.B., J.G. Henikoff, and S. Henikoff, *Genome-wide kinetics of nucleosome turnover determined by metabolic labeling of histones*. Science, 2010. **328**(5982): p. 1161-4.
16. Wollmann, H., et al., *Dynamic deposition of histone variant H3.3 accompanies developmental remodeling of the Arabidopsis transcriptome*. PLoS Genet, 2012. **8**(5): p. e1002658.
17. Rufiange, A., et al., *Genome-wide replication-independent histone H3 exchange occurs predominantly at promoters and implicates H3 K56 acetylation and Asf1*. Mol Cell, 2007. **27**(3): p. 393-405.

18.   Fal, K., et al., *Lysine 27 of histone H3.3 is a fine modulator of developmental gene expression and stands as an epigenetic checkpoint for lignin biosynthesis in Arabidopsis.* New Phytol, 2023. **238**(3): p. 1085-1100.

19.   Hodl, M. and K. Basler, *Transcription in the absence of histone H3.3.* Curr Biol, 2009. **19**(14): p. 1221-6.

20.   Yang, Y., et al., *HIRA complex presets transcriptional potential through coordinating depositions of the histone variants H3.3 and H2A.Z on the poised genes in mESCs.* Nucleic Acids Res, 2022. **50**(1): p. 191-206.

21.   Banaszynski, L.A., et al., *Hira-dependent histone H3.3 deposition facilitates PRC2 recruitment at developmental loci in ES cells.* Cell, 2013. **155**(1): p. 107-20.

22.   Wollmann, H., et al., *The histone H3 variant H3.3 regulates gene body DNA methylation in Arabidopsis thaliana.* Genome Biol, 2017. **18**(1): p. 94.

23.   Deaton, A.M., et al., *Enhancer regions show high histone H3.3 turnover that changes during differentiation.* Elife, 2016. **5**.

24.   Jamai, A., R.M. Imoberdorf, and M. Strubin, *Continuous histone H2B and transcription-dependent histone H3 exchange in yeast cells outside of replication.* Mol Cell, 2007. **25**(3): p. 345-55.

25.   Nguyen, N.H., N.T. Vu, and J.J. Cheong, *Transcriptional Stress Memory and Transgenerational Inheritance of Drought Tolerance in Plants.* Int J Mol Sci, 2022. **23**(21).

26.   Oberkofler, V. and I. Baurle, *Inducible epigenome editing probes for the role of histone H3K4 methylation in Arabidopsis heat stress memory.* Plant Physiol, 2022. **189**(2): p. 703-714.

27.   Pratx, L., et al., *Histone retention preserves epigenetic marks during heat stress-induced transcriptional memory in plants.* EMBO J, 2023. **42**(24): p. e113595.

28.   Chory, E.J., et al., *Nucleosome Turnover Regulates Histone Methylation Patterns over the Genome.* Mol Cell, 2019. **73**(1): p. 61-72 e3.

29.   Murawska, M., et al., *The histone chaperone FACT facilitates heterochromatin spreading by regulating histone turnover and H3K9 methylation states.* Cell Rep, 2021. **37**(5): p. 109944.

30.   Zhao, T., et al., *Histone H3.3 deposition in seed is essential for the post-embryonic developmental competence in Arabidopsis.* Nat Commun, 2022. **13**(1): p. 7728.

31.   Luo, Y.X., et al., *A plant-specific SWR1 chromatin-remodeling complex couples histone H2A.Z deposition with nucleosome sliding.* EMBO J, 2020. **39**(7): p. e102008.

32.   Zhao, L., et al., *Integrative analysis of reference epigenomes in 20 rice varieties.* Nat Commun, 2020. **11**(1): p. 2658.

33.   Lin, J.S., et al., *MicroRNA160 Modulates Plant Development and Heat Shock Protein Gene Expression to Mediate Heat Tolerance in Arabidopsis.* Front Plant Sci, 2018. **9**: p. 68.

34.   Silver, B.D., et al., *Differences in transcription initiation directionality underlie distinctions between plants and animals in chromatin modification patterns at genes and cis-regulatory elements.* G3 (Bethesda), 2024. **14**(3).

35.   Hani, S., et al., *Live single-cell transcriptional dynamics via RNA labelling during the phosphate response in plants.* Nat Plants, 2021. **7**(8): p. 1050-1064.

36.   Thibaud, M.C., et al., *Dissection of local and systemic transcriptional responses to phosphate starvation in Arabidopsis*. Plant J, 2010. **64**(5): p. 775-89.

37.   Hou, X., et al., *A near-complete assembly of an Arabidopsis thaliana genome*. Mol Plant, 2022. **15**(8): p. 1247-1250.

38.   Langmead, B. and S.L. Salzberg, *Fast gapped-read alignment with Bowtie 2*. Nature Methods, 2012. **9**(4): p. 357-359.

39.   Li, H., et al., *The Sequence Alignment/Map format and SAMtools*. Bioinformatics, 2009. **25**(16): p. 2078-9.

40.   Ramírez, F., et al., *deepTools: a flexible platform for exploring deep-sequencing data*. Nucleic Acids Res, 2014. **42**(Web Server issue): p. W187-91.

41.   Stempor, P. and J. Ahringer, *SeqPlots - Interactive software for exploratory data analyses, pattern discovery and visualization in genomics*. Wellcome Open Res, 2016. **1**: p. 14.

Chapter 4

Discussion

Chapter 4: Discussion

In summation, we have identified key differences in transcriptional regulation between plants and animals. The constrained geometry of the plant genome, combined with selective pressure to avoid double-stranded RNA production results in strict regulation of RNA Polymerase II direction in the sense direction at both transcription start sites (TSSs) and cis-regulatory elements (CREs). This fundamental difference in transcription directionality underscores the plant genome's organizational uniqueness; compared to animal models, which have been shown to have long-range enhancer interactions, plant CREs tend to be within 2kb upstream of their target gene [1].

Moreover, this pattern of unidirectionality does not seem to be a product of the relatively small size of the Arabidopsis genome (125 Mb); the same unidirectional pattern was observed in Maize (2,365 Mb) [2]. This suggests that this pattern of RNA Pol II regulation is specific to flowering plants as a whole. Indeed, we confirmed the same pattern in soybean (*Gylcine max)*, which is evolutionarily close to Arabidopsis; maize, however, is more distant evolutionarily speaking.

What properties of the plant genome could prevent production of reverse strand transcripts? First, unlike animals, plants contain RNA-directed DNA methylation pathways, unlike animals. Errant production of double-stranded RNAs could be greatly disruptive to the plant epigenome, thus necessitating strong selective pressure to keep RNAs tightly regulated in plants. Second, while mammalian genomes have large topologically associated domains, the plant genome is organized into small, local gene loops, where the 5' and 3' ends of a gene directly interact [3, 4]. The constrained geometry of the plant genome has been suggested to eliminate bidirectional transcription, forcing Pol II to transcribe in the sense direction alone [5].

Overall, the findings in chapter 1 highlight key distinctions in transcriptional regulation between plants and animals. Moreover, our data suggest that differences in transcriptional directionality underlie the disparities observed in chromatin modification patterns in plants and animals.

These findings underscore how little we truly understand about plant transcriptional dynamics, compared to animals, and moreover, the need for a system to measure chromatin dynamics in plants. We have developed an estradiol-inducible GFP-tagged H3.3 system for measuring turnover of H3.3 and by proxy, the entire nucleosome; in order to eject and replace H3 with any H3 variant, the entire nucleosome must be disassembled and reassembled, due to the positioning of H3 in the nucleosome [6]. Transcription is one such process that disrupts the nucleosome and can drive complete turnover. Indeed, our results identify a positive correlation between H3.3-GFP incorporation and gene transcriptional activity, suggesting greater rates of turnover at these genomic loci.

We seek to understand the critical role histone variants play in shaping the chromatin landscape, specifically the preservation of epigenetic marks. A useful tool for studying epigenetic maintenance is environmental stress. Stress-induced transcriptional memory involves the activation of stress-related genes often accompanied by H3K4 hypermethylation, suggesting a model in which stress-induced epigenetic modulation extends the duration of active transcription [7, 8]. This begs the question: when transcription is activated at these sites, how does turnover of H3.3 affect the maintenance of these stress-induced epigenetic changes? Baurle et al. [7] utilized a heat-shock (HS) inducible system to measure turnover of H3.3-GFP and found that histone turnover rates were lower at heat-shock (HS) memory genes, compared to non-memory genes; moreover, H3K4me3-marked histones were retained, suggesting that stress-

induced hypermethylation is maintained by retention of the modified histones rather than replacement of existing H3 with new H3.3.

This relationship harkens back to previously described H3.3 knockdown experiments, which showed dysregulation of environmental response genes specifically, further supporting the relationship between H3.3 and transcription activation [9]. In the same vein, H3K4me3 is a marker of recent transcriptional activity (i.e. activation); ergo, the retention of H3.3, and as a result, H3K4me3, at response genes further indicates a role in turnover of H3.3 and gene activation [7].

While the ability to measure histone turnover is certainly a useful tool, there are caveats to usage of the inducible H3.3-GFP system. First, while we are able to measure an increase in H3.3-GFP incorporation after induction during a time course, the actual "rate" of H3.3 turnover is more so of an arbitrary measurement. Previous studies using inducible methods to measure histone turnover have employed an exponential decay model to report the mean lifetime of histones as representing a single full cycle of turnover [7, 10]. This allows for normalization of all time points to the start of induction and modeling of the kinetics of H3.3-GFP turnover. This also addresses a second caveat about this technique. By the nature of being an inducible system, we are flooding the plant with H3.3, most likely at greater volumes than is naturally occurring under normal conditions; this can lead to somewhat of an issue in modeling true biological behaviors of nucleosome turnover. Hence, the environmental decay function can provide a better sense of histone decay, relative to the time points measured during induction.

Taken together, our results highlight fundamental differences and similarities in transcription activation and chromatin dynamics between plants and animals. During a cross-species meta-analysis of transcription-associated PTMs, accessible chromatin and nascent RNA

data, we observed that transcription in plants is highly regulated and strictly unidirectional at both the transcription start site and potential enhancer elements. In contrast, animals were observed to have bidirectional transcription at these sites. These stark differences in chromatin modifications and RNA Polymerase II activity suggest that there are potential differences in chromatin dynamics during transcription activation.

Overall, the results described here provide insights into the fundamental differences between plant and animal transcriptional regulation, while providing a system to further investigate plant transcriptional activity in the greater chromatin context. Future experiments are needed to ascertain epigenetic and transcriptional changes that accompany nucleosome turnover in response to environmental stimuli. This will ultimately crystallize the relationship between histone retention and the inheritance of stable epigenetic states.

References

1.   Maher, K.A., et al., *Profiling of Accessible Chromatin Regions across Multiple Plant Species and Cell Types Reveals Common Gene Regulatory Principles and New Control Modules*. Plant Cell, 2018. **30**(1): p. 15-36.

2.   Haberer, G., et al., *Structure and architecture of the maize genome*. Plant Physiol, 2005. **139**(4): p. 1612-24.

3.   Liu, C., et al., *Genome-wide analysis of chromatin packing in Arabidopsis thaliana at single-gene resolution*. Genome Res, 2016. **26**(8): p. 1057-68.

4.   Lee, H. and P.J. Seo, *Accessible gene borders establish a core structural unit for chromatin architecture in Arabidopsis*. Nucleic Acids Res, 2023. **51**(19): p. 10261-10277.

5.   Tan-Wong, S.M., et al., *Gene loops enhance transcriptional directionality*. Science, 2012. **338**(6107): p. 671-5.

6.   Robert, F. and C. Jeronimo, *Transcription-coupled nucleosome assembly*. Trends Biochem Sci, 2023. **48**(11): p. 978-992.

7.   Pratx, L., et al., *Histone retention preserves epigenetic marks during heat stress-induced transcriptional memory in plants*. EMBO J, 2023. **42**(24): p. e113595.

8.   Oberkofler, V. and I. Baurle, *Inducible epigenome editing probes for the role of histone H3K4 methylation in Arabidopsis heat stress memory*. Plant Physiol, 2022. **189**(2): p. 703-714.

9.   Wollmann, H., et al., *The histone H3 variant H3.3 regulates gene body DNA methylation in Arabidopsis thaliana*. Genome Biol, 2017. **18**(1): p. 94.

10.   Deal, R.B., J.G. Henikoff, and S. Henikoff, *Genome-wide kinetics of nucleosome turnover determined by metabolic labeling of histones*. Science, 2010. **328**(5982): p. 1161-4.