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:

_____ _____

Kelsey A. Maher Date

APPLICATIONS OF NEXT-GENERATION SEQUENCING STRATEGIES FOR THE IDENTIFICATION AND CHARACTERIZATION OF ENHANCERS IN PLANTS

By
Kelsey A. Maher
Doctor of Philosophy

Graduate Division of Biological and Biomedical Science
Biochemistry, Cell, and Developmental Biology

Roger B. Deal, Ph.D.
Advisor

Anita H. Corbett, Ph.D.
Committee Member

David J. Katz, Ph.D.
Committee Member

William G. Kelly, Ph.D.
Committee Member

Kenneth H. Moberg, Ph.D.
Committee Member

Accepted:

Lisa A. Tedesco, Ph.D.
Dean of the James T. Laney School of Graduate Studies

Date

APPLICATIONS OF NEXT-GENERATION SEQUENCING STRATEGIES FOR THE IDENTIFICATION AND CHARACTERIZATION OF ENHANCERS IN PLANTS

By

Kelsey A. Maher
B.S. Boston College, 2014

Advisor: Roger B. 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 fulfilment of the requirements for the degree of
Doctor of Philosophy
In Biochemistry, Cell, and Developmental Biology
2020

Abstract

APPLICATIONS OF NEXT-GENERATION SEQUENCING STRATEGIES FOR THE IDENTIFICATION AND CHARACTERIZATION OF ENHANCERS IN PLANTS

By Kelsey A. Maher

The transcriptional regulatory structure of plant genomes remains poorly defined relative to animals. It is unclear how many *cis*-regulatory elements exist, where these elements lie relative to promoters, and how these features are conserved across plant species. This is due in part to the challenges presented by plant tissues, including a resilient cell wall, an abundance of extra-nuclear DNA, and the inability to maintain single-cell types via cell culture, making them less than ideal candidates for the latest next-generation sequencing technologies. Here we present two approaches to isolate nuclei with high purity from plant tissues for input into Assay for Transposase-Accessible Chromatin with high-throughput sequencing (ATAC-seq). We used this method to delineate open chromatin regions and transcription factor (TF) binding sites across the *Arabidopsis thaliana*, *Medicago truncatula*, *Solanum lycopersicum*, and *Oryza sativa* genomes. We find that the majority of open chromatin regions lie within 3 kb upstream of a transcription start site, and that TF-gene networks in the root tips are broadly conserved between the four species. Comparative ATAC-seq profiling of *Arabidopsis* root hair and non-hair cell types revealed extensive similarity on a global level, while TF motif analysis of differentially accessible chromatin regions identified a MYB-driven regulatory module unique to the hair cell's fate and function. Finally, we generate single-cell type chromatin immunoprecipitation (ChIP-seq) datasets for the four histone modifications conserved across animal enhancers. Combined with our single-cell type chromatin accessibility data and available nascent transcript data, we compare the conservation of enhancer epigenetic characteristics between *Arabidopsis thaliana*, *Homo sapiens*, and *Drosophila melanogaster*. While animal promoters and enhancers show characteristic bimodal histone mark deposition and bidirectional transcription, this analysis revealed that *Arabidopsis thaliana* promoters and enhancers exclusively exhibit histone mark deposition and transcription in the sense direction, which may speak to a fundamental difference between transcriptional initiation in the plant and animal kingdoms. Together this work has interrogated the location, quantity, and characteristics of putative enhancers in plants on multiple levels, in multiple species, and in multiple cell types, and has generated new methodologies for the continued investigation of *cis*-regulatory elements in plants in the future.

APPLICATIONS OF NEXT-GENERATION SEQUENCING STRATEGIES FOR THE IDENTIFICATION AND CHARACTERIZATION OF ENHANCERS IN PLANTS

By

Kelsey A. Maher
B.S. Boston College, 2014

Advisor: Roger B. Deal, Ph.D.

A dissertation submitted to the Faculty of the
James T. Laney School of Graduate Studies of Emory University
In partial fulfilment of the requirements for the degree of
Doctor of Philosophy
In Biochemistry, Cell, and Developmental Biology
2020

Acknowledgements

I would like to thank the faculty that have offered their effort and time to be a part of my committee: Bill Kelly, Ken Moberg, Dave Katz, Anita Corbett, and former members Paula Vertino and Xiaodong Cheng. I would also like to thank the endless support and optimism of everyone in the Biology Department, especially the staff, who have helped me solve problems both extraordinary and mundane during my time at Emory. To the members of the Deal Lab, past and present: Shannon, Paja, Kris, Dylan, Ellen, Maryam, Tom, and all the undergraduates I've had the pleasure to work with – you were the greatest lab family I could have asked for; your laughter, your kindness, and your passion for the work will be dearly missed. Most especially Marko – you have helped me out of problems time and time again, from debugging code to navigating conventions to whatever else I threw your way. You are truly a one-of-a-kind person, and I thank you from the bottom of my heart for everything you've done. Roger, your enthusiasm is infectious, and I wish every scientist embodied your boundless passion for discovery. I could never have imagined a better mentor. You have always treated me like a person first, and an employee second, and that has made all the difference. Thank you for believing in me, especially when I didn't believe in myself.

I would like to thank my cohort, Regan, Emily, Ed, Amanda, Sabrina, and David, and the friends I've made at BCDB and Emory at large – you were always there to rejoice in the good times, commiserate in the bad times, and muddle our way through the rest without losing our sense of humor. I would like to take a moment to extend a very special thanks to Sabrina; I know could not have made it through graduate school without you. I am so grateful that our paths crossed and I know that they will continue to do so as life-long friends. I would like to thank my friends from Boston College and my childhood friends from East Lyme – our enduring ties have been a salve, a boon, and delight through the ups and downs of life. To my brother, Justin, and to my parents, whose tireless outpouring of love and support has buoyed me in pursuing my goals. Thank you for all you have done for me, now and always. And finally, I would like to thank my fiancé Josh. You have been with me since the very beginning of this grad school journey, as a friend and now as my husband-to-be, and I know you will be by my side for all of my tomorrows. Thank you for you love, your encouragement, your everything. I could not have done it without you, my love.

Table of Contents

CHAPTER 1: INTRODUCTION

Gene Expression Regulation Begins with DNA

The regulation of transcription is the most fundamental and versatile tool in the arsenal of living organisms. The ability to control gene expression impacts all functions of living creatures, allowing bacteria to defend against microscopic assailants and providing plants and animals the ability to respond in real time to ever-changing environmental demands. Even more remarkably, it allows a single-celled zygote to develop into a mature adult, a being comprised of hundreds of distinct cell types, each with different form and function but nonetheless identical genetic information.

The regulation of these processes begins on the molecular level with our genetic information, in the form of double-stranded molecules of deoxyribonucleic acid (DNA). Comprised of a sugar-phosphate backbone and nucleotide bases (adenosine (A), thymine (T), cytosine (C), and guanine (G)), the specific bonding of bases between two DNA strands (A with T, C with G) stores information that can be faithfully maintained through replication, allowing it to be passed on to future offspring. The diploid human genome is comprised of over 6 billion nucleotide base pairs (bp), resulting in strands of DNA that stretch ~2 meters long when laid end to end. This composite molecule – longer than the average person is tall – is a tremendous amount of information to fit into a cell ten times smaller than the period at the end of this sentence. How can such an extraordinary feat be accomplished?

DNA is Packaged into Chromatin

Much like thread on a spool, the genome is packaged into units called 'nucleosomes' made of ~147 base pairs (bp) of DNA wrapped around an octamer of proteins called 'histones'(Kornberg 1977, Luger, Mader et al. 2000). Histone proteins contain both a 'globular domain', which comprises the core of the nucleosome, and a 'tail domain', which extends from the periphery. The degree to how tightly the nucleosome associates with the DNA is determined in part by which chemical modifications are added to

the amino acids of the histone proteins; most commonly, the residues in the histone tail domain are modified, though amino acids in the globular domain can be altered as well. By virtue of their highly basic amino acid composition, histones are positively charged proteins, making them electrostatically attracted to DNA, whose phosphate backbone carries a negative charge. Modifications can decrease the histone's positive charge (e.g. acetylation, phosphorylation), causing a decreased attraction between the nucleosome and the DNA. Additionally, these marks as well as modifications which do not alter charge (e.g. methylation) can impact attraction by recruiting protein factors referred to as 'readers' to the site, which either contain regulatory domains themselves or in turn recruit additional regulators.

Furthermore, modifications can generate or inhibit binding sites for other *trans*-acting protein factors. These chemical groups are added and removed after the histones have been translated and processed, making them posttranslational modifications (PTMs) to the histone proteins.

Each standard, canonical nucleosome is comprised of two H2A-H2B histone dimers, and a dimer of histone H3-H4 dimers (Arents, Burlingame et al. 1991). Just as histones can be altered after the fact by post-translational chemical modifications (PTMs), variant forms of histones are encoded in the genome and can expressed for a range of purposes. Core histones (H2A, H2B, H3, H4) are generated during S phase of the cell cycle when extra nucleosomes are required to package the newly synthesized duplicate DNA for cell division (Marzluff and Duronio 2002). Histone variants are expressed independent of replication, and are incorporated into nucleosomes in the place of their core histone counterparts (Marzluff, Gongidi et al. 2002), either passively or by the activity of ATP-dependent chromatin remodeling complexes (Clapier and Cairns 2009). Common variants include CENP-A, a H3 variant crucial for establishing and maintaining centromere identify (Palmer, O'Day et al. 1991), and H2A.X, involved in repair of DNA at double-strand breaks (Lowndes and Toh 2005, Morrison and Shen 2009). Ultimately, by packaging DNA around histone octamers the large, unwieldly linear genome is condensed into the more compact 'beads on a string' nucleosome array approximately 11 nanometers wide (Olins and Olins 1974).

It should be noted that nucleosomes are not spaced in a regular distribution across the genome. Nucleosomes are actively slid, condensed, disassembled, and even evicted from their positions by ATP-

dependent chromatin remodeling complexes to allow – or bar – free-floating *trans*-acting factors access to the underlying DNA sequence (Clapier and Cairns 2009, Luger, Dechassa et al. 2012). As different sequences are required at different stages of the life cycle, the spacing and compaction of nucleosomes can change both subtly and dramatically throughout the life of the organism.

Higher orders of compaction are utilized to condense the genome even further. With the addition of linker histone H1 (Robinson and Rhodes 2006) and proteins including HP1 (Canzio, Chang et al. 2011) and Polycomb (Francis, Kingston et al. 2004), nucleosome arrays are folded into a ~30 nm chromatin fiber (Song, Chen et al. 2014). This fiber is compacted into chromatin domains ranging 300 nm – 700 nm in diameter, with regions secured to the nuclear periphery by nuclear lamins (Amendola and van Steensel 2014). These Lamina-Associated Domains (LADs) along with Topological-Associated Domains (TADs) help organize the epigenome into interactive, functional units by facilitating long-range chromatin interactions (Gonzalez-Sandoval and Gasser 2016).

Finally, in preparation for specific stages of the cell cycle, chromatin domains are further condensed and organized into mitotic chromosomes. These remarkable structures represent a 10,000-fold compaction of the original linear DNA, and allow for the easy partitioning of genetic material between daughter cells during cellular replication. Several protein factors and complexes, including topoisomerase II, condensin, cohesin, and hyperphosphorylated linker histone H1 are required to achieve this incredible degree of compaction (Nasmyth and Haering 2009, Hirano 2012) (**Figure 1.1**).

Heterochromatin

The compressed form of chromatin, referred to as heterochromatin, encompasses a wide variety of condensation states with a multitude of purposes in the cell. During the radical process of cellular replication, where the genetic information is physically manipulated, heterochromatin structures protect regions which endure high levels of mechanical stress such as telomeres – the ends of chromosomes – and centromeres – where the mitotic machinery attaches to each chromosome (Buhler and Gasser 2009). Compaction also severely inhibits transcription, as the nucleotide base sequence itself must be accessible

to be ‘read’ by *trans*-acting protein factors in order to generate RNA and later create functional protein. Thus, most genetic sequences caught in a region of constitutive heterochromatin are transcriptionally silenced as they are inaccessible to *trans*-factors. This phenomenon is often used to the cell’s advantage. Long regions of genomic repeats, usually indicative of invasive viral DNA, are condensed into constitutive heterochromatic regions to prevent their activation (Pikaard and Mittelsten Scheid 2014, Strome, Kelly et al. 2014, Allshire and Ekwall 2015, Martienssen and Moazed 2015). This is achieved by a variety of mechanisms, including deacetylating histones, trimethylating lysine nine of histone 3 (H3K9me3), recruiting heterochromatin architectural proteins such as HP1, and establishing cytosine methylation on the surrounding DNA. While constitutive heterochromatic regions remain compacted throughout the cell cycle, facultative heterochromatin can be reversibly established during specific stages of the cell cycle or development. Rather than H3K9me3, regions of facultative heterochromatin are marked by H3K27me3, which in turn recruits a variety of factors, including the chromatin-compacting Polycomb Repressive Complexes (PRCs). This pathway is used to stably and reversibly silence genes and regulatory elements, allowing the cell to deactivate unused developmental pathways as needed.

Euchromatin Contains Genes, Promoters, and Cis-Regulatory Elements

Much of the epigenome exists in a non-condensed state, called euchromatin, comprised largely of gene-rich regions. Euchromatin lacks the three-dimensional compaction created by the architectural proteins present in heterochromatin. Specialized transcription factors, called ‘pioneer TFs’, have the unique ability to associate with their target DNA binding sequence, or ‘transcription factor motif’, regardless of whether the motif is freely available or packaged into a more condensed chromatin structure (Sherwood, Hashimoto et al. 2014, Soufi, Garcia et al. 2015). Working in concert with chromatin remodeling complexes, these pioneer transcription factors can convert regions of heterochromatin into euchromatin. While every element in euchromatic regions is not active simultaneously, the decondensed chromatin structure means that these sequences are accessible to *trans*-acting machinery and may become more easily activated.

Accessible chromatin regions are historically referred to as nucleosome-free or nucleosome-depleted regions (NFRs, NDRs) due to their lack of signal from chromatin immunoprecipitation with sequencing (ChIP-seq) assays. It has become clear that, rather than these regions being devoid of nucleosomes entirely, this absence of ChIP-seq signal is due to the high rate of turnover the nucleosomes experience in these regions. As nucleosomes compete for binding sites with DNA-binding transcription factors, histones are not associated with the region for a long enough duration to be registered by the assay. Additionally, many NDRs/NFRs contain histone variants that are susceptible to the high salt concentrations used in many ChIP-seq procedures, making them vulnerable to eviction during the assay itself (Jin and Felsenfeld 2007, Henikoff, Henikoff et al. 2009, Jin, Zang et al. 2009). As a result, a key characteristic of regulatory elements in euchromatin is their hypersensitivity to chromatin accessibility-probing assays, such as DNase-seq (Boyle, Davis et al. 2008), MNase-seq (Yuan, Liu et al. 2005), and ATAC-seq (Buenrostro, Giresi et al. 2013, Buenrostro, Wu et al. 2015) (**Figure 1.2**). Other nucleosomes in euchromatin, especially the well-positioned nucleosomes flanking the borders of regulatory elements, are often marked with a wide range of histone PTMs. These modifications can denote classes of genomic elements (i.e. H3K36me on gene bodies) as well as the level of activity of the region (H3K4me3 on active promoters, H3K4me3/H3K27me3 on a bivalent/poised promoters (Bernstein, Mikkelsen et al. 2006)).

A primary feature of euchromatin is that it houses genes, sequences in DNA that code for RNA and subsequently, in many cases, protein products. Gene sequences are comprised of nucleotide bases which – after being transcribed into an intermediary RNA molecule – may be 'read' sequentially by a ribosome in groups of threes. Each trio of bases codes for a specific amino acid, allowing the genetic DNA sequence to be converted into functional protein. Gene bodies begin with the transcription start site (TSS), where the cellular machinery will begin the process of transcribing the DNA sequence of the gene body into RNA. Genes are regulated by at least one promoter, located ~50-200 bp upstream of the gene's TSS, which can be identified by a variety of conserved sequence combinations, including CpG islands and TATA boxes. Promoters recruit transcriptional machinery, including general transcription factors and RNA Polymerase II, to form the Pre-Initiation Complex (PIC) at the TSS in preparation for genic transcription. In this way,

promoters allow genes to be expressed at a basal level; however, to exert any finesse over the degree, timing, and location of transcriptional output, organisms need the input of *cis*-regulatory elements (CREs).

Cis-regulatory elements (CREs) act by largely the same basic mechanism, by providing a binding platform for *trans*-acting factors in the nucleoplasm (Sims, Belotserkovskaya et al. 2004, Voss and Hager 2014). These platforms are comprised of a collection of transcription factor (TF) binding motifs, short sequences of DNA about 6-12 base pairs long. Because TF binding motifs are often degenerate, meaning a single protein factor can bind to a variety of similar but non-identical sequences, TFs are recruited to regulatory regions by more than simply their baseline affinity for the DNA motif. Numerous forces work in concert to allow TFs to bind in high concentrations to regulatory elements, including the opening of new binding sites by chromatin remodeling, the inhibition of nucleosome repositioning, the bending of the DNA molecule into a more favorable architecture, and the affinity of TFs for neighboring TFs or co-factors (Spitz and Furlong 2012). However, despite the commonality of this basic mechanism, different categories of *cis*-regulatory elements fulfill very distinct functions in the cell. Silencer CREs provide a binding platform for repressor factors and in turn suppress the activity of target promoters. This silencing can be achieved by the element's direct interference with the basal transcriptional machinery, or by competing with enhancer elements, but ultimately results in the reduction of transcriptional output of the target genes (Ayer and Benyajati 1990, Petrykowska, Vockley et al. 2008, Vokes, Ji et al. 2008). Insulator CREs have one of two main functions: first, to act as a boundary to prevent the spread of heterochromatin into euchromatic areas, or second, to restrict promoter-stimulating activity from affecting regions beyond the local chromatin domain (Gaszner and Felsenfeld 2006). Enhancers, a third type of *cis*-regulatory element, recruit transcription factors to activate a target promoter, resulting in the increased transcriptional output of the target gene. Enhancers are perhaps the best characterized of the *cis*-regulatory elements to date, and their identification and characterization, especially in plant genomes, are a focus of this dissertation.

Enhancers

Enhancer elements are a highly conserved type of genetic control mechanism, and have been found in a diverse array of organisms, from complex eukaryotes (Schwaiger, Schonauer et al. 2014, Villar, Berthelot et al. 2015, Zhu, Zhang et al. 2015, Weber, Zicola et al. 2016) to rudimentary bacteria (Xu and Hoover 2001) and viruses (Berg, Popovic et al. 1984). On a molecular scale, genetic enhancers consist of DNA sequences ranging between tens to hundreds of base pairs in length. These sequences are comprised of a modular collection of transcription factor binding motifs which in turn act as an assembly platform for *trans*-acting factors (Lee and Young 2000, Spitz and Furlong 2012). Similar to promoters, enhancers act as a binding site for sequence-specific transcription factors (TFs), general TFs, and co-factors, which in turn physically associate with RNA Polymerase II to form the Pre-Initiation Complex (PIC) (Sainsbury, Bernecky et al. 2015). In turn, larger molecular machinery is recruited. This includes the protein complex scaffold, Mediator (Ebmeier and Taatjes 2010, Kagey, Newman et al. 2010, Poss, Ebmeier et al. 2013), and cohesin (Kagey, Newman et al. 2010, Schmidt, Schwalie et al. 2010, Faure, Schmidt et al. 2012), which stabilizes the enhancer-promoter interaction. Furthermore, nucleosome remodelers and histone modifying proteins such as the histone acetyltransferase CPB/p300 (Vernimmen and Bickmore 2015) are recruited to make alterations to the surrounding epigenomic landscape.

Ultimately, the *cis*-regulatory element and its associated *trans*-factors will maneuver in 3D space to interact with a target promoter (Carter, Chakalova et al. 2002, Tolhuis, Palstra et al. 2002, Ghavi-Helm, Klein et al. 2014). The close association of these elements and their accompanying factors enriches the local microenvironment (Lemon and Tjian 2000, Kulaeva, Nizovtseva et al. 2012) for activating TFs and assembled transcriptional machinery, and stimulates the transcription of the target gene. The RNA molecule that is produced will then be processed to generate a functional protein, or go on to have biochemical activity itself. By serving as a binding platform for a variety of factors and complexes, enhancers act as a crucial mechanism to integrate a myriad of cellular signals into meaningful transcriptional output. This allows the cell and the organism at large to respond dynamically to both internal cues (i.e. developmental signals) and external cues (i.e. environmental stimuli) alike, permitting it to maintain precise spatiotemporal control over its gene expression.

Enhancers are Challenging to Study

Enhancers provide a crucial and necessary function to the cell, but these elements have historically presented an enormous challenge to study. Promoters can be readily identified by a suite of characteristics, from conserved sequence motifs (TATA boxes and CpG islands), to their proximal location upstream of gene bodies, and even by their preference to function in the forward orientation with respect to their target gene (Kim and Shiekhattar 2015). Enhancers, in contrast, have startling few commonalities between elements. Beyond the DNA-binding motifs of individual TFs, this class of *cis*-regulatory sequences lacks any sort of overarching sequence conservation (Villar, Berthelot et al. 2015). Enhancers have been found to be able to act in both 'forward' and 'reverse' orientations with regard to their target promoter, and functional elements have been found to be abundant in both the genic and intergenic regions of the genome. Additionally, while evidence suggests that enhancers preferentially regulate the most proximal promoter (Heintzman, Hon et al. 2009, Anders and Huber 2010, Creyghton, Cheng et al. 2010), other elements have been characterized up to tens of thousands of base pairs away from their target promoters, or even on entirely separate chromosomes (Lettice, Heaney et al. 2003, Kleinjan and van Heyningen 2005, Ong and Corces 2011). Taken together, this remarkably permissive profile leaves very little concrete criteria on which to positively identify an enhancer element in a discriminating manner, making the study of these elements uniquely challenging.

Animal Enhancers have a Unique Set of Secondary Characteristics

Over the past decade and a half, the combined efforts of the Encyclopedia of DNA Elements (ENCODE) Project (2004, 2012) and modENCODE (Boley, Wan et al. 2014) have vastly accelerated the characterization of regulatory elements in humans and animal models. Beginning in 2003 (Birney, Stamatoyannopoulos et al. 2007), the ENCODE Project represents a massive consortium effort to analyze thousands of experimental and computational datasets in order to better understand the functional elements of the metazoan genome. While originally focused on humans, the database has expanded to include over

13,000 datasets for human, mouse, *Drosophila*, and *Caenorhabditis elegans*, encompassing a variety of tissues, cell lines, and physiological states (Davis, Hitz et al. 2018). Through these analyses and others, the epigenetic qualities of enhancer elements in animal species came into focus.

First and foremost, regulatory elements are binding platforms for transcription factors and regulatory machinery. Techniques that probe the relative accessibility of chromatin to *trans*-acting factors, especially high-throughput next-generation sequencing technologies including MNase-seq (Yuan, Liu et al. 2005), DNase-seq (Keene, Corces et al. 1981, McGhee, Wood et al. 1981, Boyle, Davis et al. 2008), and ATAC-seq (Buenrostro, Giresi et al. 2013, Buenrostro, Wu et al. 2015) (**Figure 1.2**), reveal that enhancers and other regulatory regions of DNA preferentially exist in areas of open chromatin (Tsompana and Buck 2014, Jiang 2015). ChIP-seq assays have revealed that a variety of *trans*-acting factors, including paused RNA Polymerase II, Mediator, CPB/p300, and others remain associated with these accessible regions consistently enough to be used as identifying features of enhancers in their own right (Zentner and Scacheri 2012, Heinz, Romanoski et al. 2015). The DNA of active enhancers is often hypomethylated, similar to that of active promoter elements (Angeloni and Bogdanovic 2019). Also similar to active promoters, active enhancers have been shown to produce transcripts, called 'enhancer RNAs' or 'eRNAs'. This species of non-coding RNA varies widely, with some molecules being short and transcribed bi-directionally, and others being long, uni-directional, and polyadenylated. While the exact function of eRNAs has yet to be determined, speculations range from the small RNAs serving a TF-recruiting mechanism to being a mere by-product of an over-active RNA Polymerase II (Lai and Shiekhattar 2014).

Enhancers are enriched for nucleosomes with histone variants H2A.Z and H3.3, whose instability is proposed to contribute to the hyperaccessibility of enhancers (Barski, Cuddapah et al. 2007, Jin and Felsenfeld 2007, Wang, Zang et al. 2008, Henikoff, Henikoff et al. 2009, Jin, Zang et al. 2009). Furthermore, through the use of genome-wide chromatin immunoprecipitation with high-throughput sequencing (ChIP-seq), a variety of histone posttranslational modifications (PTMs) have been found to be associated with the well-positioned nucleosomes that flank enhancers (Wang, Zang et al. 2008, Hawkins, Hon et al. 2010, Ernst, Kheradpour et al. 2011, Zentner, Tesar et al. 2011, Bonn, Zinzen et al. 2012). These

include H3K9me1 (Barski, Cuddapah et al. 2007, Wang, Zang et al. 2008), H3K18ac (Wang, Zang et al. 2008), H3K9ac and H3K14ac (Roh, Cuddapah et al. 2005, Roh, Wei et al. 2007). Among these studies, a single set of marks has emerged as the most highly conserved across enhancers in a variety of cell types and species: H3K27ac, H3K27me3, H3K4me1, and H3K4me3. As several of these modifications have also been reported to have overlap with promoters, a higher ratio of H3K4me1/H3K4me3 enrichment has been used as a guideline to distinguish enhancers from their gene-proximal counterparts (Heintzman, Stuart et al. 2007, Heintzman, Hon et al. 2009, Kim and Shiekhattar 2015). These four histone PTMs rarely coincide on a single element at once; rather, combinations of these marks denote different activity states of enhancers. 'Active' enhancers are marked by H3K4me1 and H3K27ac and have high eRNA production and high transcriptional output of their target promoter(s) (Creyghton, Cheng et al. 2010). 'Poised' enhancers are characterized by H3K4me1 and H3K27me3, and low transcriptional promoter output (Rada-Iglesias, Bajpai et al. 2011, Zentner, Tesar et al. 2011). 'Intermediate' enhancers, which are believed to be in transition between these two states, are characterized by H3K4me1 alone (Zentner, Tesar et al. 2011, Zentner and Scacheri 2012). These enhancer qualities are summarized in **Figure 1.3**.

Plant Enhancer Studies are Advancing

Without a huge consortium effort paralleling that of the ENCODE Project, the progress of regulatory element characterization in plant systems has lagged decades behind the efforts made for the animal kingdom. Early research into plant enhancers dates back to the 1980s with the discovery of an enhancer of the *AB80* gene, a chlorophyll *a/b*-binding protein, in pea plants (Simpson, M et al. 1986), and continued with the characterization of the tb1 enhancer in maize (Clark, Wagler et al. 2006, Studer, Zhao et al. 2011). Up until 2014, *cis*-regulatory element (CRE) studies in the plant kingdom were limited low-throughput techniques, such as laborious promoter deletion assays, electrophoretic mobility shift assays (EMSA), and single-gene DNase-I footprinting (Timko, Kausch et al. 1985, Green, Kay et al. 1987, Valles 1991). Enhancer trapping studies in *Arabidopsis* and rice were able to identify a few tissue-specific enhancers, but the random insertion of the 'bait' minimal promoter construct across the genome made reliable

identification of interacting enhancers difficult (Wu, Li et al. 2003, Yang, Jefferson et al. 2005, McGarry and Ayre 2008, Chudalayandi 2011). As a result, fewer than two dozen bona fide enhancers had been identified across the entire plant kingdom, concentrated primarily in just three species: *Arabidopsis thaliana*, *Zea mays*, and *Pisum sativum* (Weber, Zicola et al. 2016).

Plant enhancer research finally began to keep pace with its animal counterpart with the application of next-generation sequencing techniques to Plantae genomes. Genome-wide chromatin accessibility data in the form of DNase-I hypersensitive site (DHS) mapping and micrococcal nuclease sequencing (MNase-seq) allowed for the drastic scaling up of enhancer investigation, identifying thousands of putative regulatory regions across *Arabidopsis thaliana* (Zhu, Zhang et al. 2015), *Oryza sativa* (Zhang, Wu et al. 2012, Zhang, Zhang et al. 2012, Sullivan, Arsovski et al. 2014, Zhu, Zhang et al. 2015), and *Zea mays* (Rodgers-Melnick, Vera et al. 2016). However, these studies rely on whole tissue or whole organisms as input. In line with their role as drivers of embryonic development and cell specification programs (Chatterjee and Ahituv 2017), the activity of enhancers varies drastically in regard to individual cell type (Ernst and Kellis 2010, Rada-Iglesias, Bajpai et al. 2011, Zentner, Tesar et al. 2011). As such, it is ideal to compare sequencing datasets drawn from the same cell type. This proves to be a challenge because cultured immortalized cell lines are not available in plants as they are in animals, meaning that any cell type must be directly isolated from whole, living organisms. Nonetheless, studies which examine chromatin signatures across a whole tissue or organism risk muddying enhancer activity state signals (Creyghton, Cheng et al. 2010, Rada-Iglesias, Bajpai et al. 2011) by averaging enrichment trends together across several cell types. This effect dampens regions of extreme enrichment or depletion, making genuine elements more difficult to distinguish from background noise.

Scope of the Dissertation

In **Chapter 2** of this dissertation, I present work I completed with another graduate student in the lab, Marko Bajic, to design an improved technique for identifying accessible chromatin sites genome-wide in plants with single cell-type specificity. Rather than relying on time-consuming approaches such as

DNase-seq and MNase-seq (**Figure 1.2**), we adapt Assay for Transposase Accessible Chromatin with high-throughput sequencing (ATAC-seq) for use in plant species, which can yield tagged libraries from nuclei in half an hour. Using a hyperactive Tn5 transposase pre-loaded with sequencing adapters, ATAC-seq simultaneously fragments and labels the accessible genome in a single step called 'tagmentation', making this procedure vastly quicker to execute than DNase-seq while yielding results of comparable quality. A noticeable downside of this approach, however, is the hyperactive transposase's tendency to act on non-nuclear sources of DNA in the cell, generating an abundance of reads that must be discarded before analysis. This issue is compounded further in plants which have both mitochondrial and chloroplastic DNA, making it exceptionally challenging to get an adequate sequencing depth in the nuclear genome. To overcome this challenge, we pair ATAC-seq with nuclei-purifying steps in our protocol: Isolation of Nuclei Tagged in specific Cell Types (INTACT) (Deal and Henikoff 2010) for affinity purification of single cell-types in transgenic lines, and sucrose sedimentation for general nuclei purification. Combined, our protocol produces sequencing-ready ATAC-seq libraries in less than a day with the option of single cell-type specificity, extremely low organellar reads, and a procedure that can be readily used in a variety of plant species.

In **Chapter 3**, I, Marko Bajic, and collaborators from the University of Washington, University of California Davis, and University of California Riverside apply our method of plant-based ATAC-seq to launch a cross-species comparison of accessible chromatin with the aim of uncovering putative *cis*-regulatory regions. The species of interest include *Arabidopsis thaliana*, the hallmark model plant; *Medicago truncatula*, a model legume; *Oryza sativa*, rice; and *Solanum lycopersicum*, the domesticated tomato. Though the genome size and genic density vary widely between these species, we see that consistently the majority of open chromatin sites lie within a 3 kb window upstream of a transcription start site (TSS). Within the root tips of all four species, we uncovered a common set of four transcription factors (TFs) whose binding motifs were enriched in open chromatin regions, indicating that TF-gene networks are generally conserved. In addition to a cross-species comparison, we also conducted a cross-cell lineage comparison to investigate what chromatin accessibility could reveal about the differential regulation

pathways of developmentally related cell types. I generated ATAC-seq libraries profiling *Arabidopsis* root hair and non-hair cells, which revealed a surprising level of similarity between the chromatin accessibility of the two cell types. However, on closer examination utilizing quantitative analysis of accessibility within the regions of accessible chromatin, we uncovered a MYB-driven regulatory module unique to the hair cell which appears to control both cell fate regulators and abiotic stress responses. Our analyses revealed common transcriptional regulatory principles across species, and shed light on fundamental mechanisms producing cell-type-specific transcriptomes during development.

In addition to chromatin hyperaccessibility, we aimed to determine whether additional epigenetic enhancer characteristics are conserved between plants and animals. Several qualities, such as DNA methylation, do not lend themselves to straightforward comparisons between plant and animal epigenomes. While 5-methylcytosine (5mC) is predominantly found at CG sites in animal genomes, CG, CHG, and CHH motifs all can be methylated in plants (H can represent A, T, or C) (Furner and Matzke 2011, Meyer 2011), indicating that methylation states cannot be translated 1:1. Animal promoters are usually found in methylation-free 'CpG islands', but analogous regions have proven difficult to identify in plants. Furthermore, while dynamic DNA methylation levels are used in animal models to regulate the activity of tissue-specific enhancers, this has not been found to be the case in plants (Weber, Zicola et al. 2016), suggesting that the function and regulatory mechanisms behind DNA methylation may be significantly different between the kingdoms.

Due to a higher tolerance for genome duplication, gene families in plants tend to be much more extensive than they are in animals, making it extremely challenging to identify and characterize functional homologs. Many of the most common enhancer-associated *trans*-factors in animals, such as CTCF, do not have direct orthologs in plants. Of the few subunits of Mediator that have been identified in *Arabidopsis*, most show low homology to subunits in other species, and several are in fact plant-specific (Backstrom, Elfving et al. 2007). Few ChIP-seq datasets exist for transcription factors in plants (Bolduc, Yilmaz et al. 2012, Yu, Chen et al. 2015), with most datasets having been generated from *in vitro* DNA Affinity Purification with sequencing (DAP-seq) experiments (Mathelier, Zhao et al. 2014, O'Malley, Huang et al.

2016). With the limited information available, teasing apart this homology network can be prohibitively complex for a single gene product, let alone a collection of enhancer complexes. As such, while mapping of *trans*-acting factors has been used with great success for enhancer discovery in animal species, this approach would not be easy to adapt to plants. Therefore, we chose to pursue three of the well-established enhancer criteria – chromatin accessibility, eRNA production, and histone marks – as means for enhancer discovery in plants.

In **Chapter 4** of this dissertation, I expand on the work of a former postdoctoral fellow in the lab, Dongxue Wang, Ph.D., to explore the question of histone modification conservation between plant and animal enhancers. She generated and I analyzed ChIP-seq datasets with single cell-type specificity for the highly conserved set of animal enhancer histone PTMs H3K27ac, H3K27me3, H3K4me1, and H3K4me3, in the model plant *Arabidopsis thaliana*. We utilized root epidermal non-hair cells, one of the same cell types explored in Chapter 3, making our libraries highly comparable to our previous datasets. Combined with our single cell-type non-hair cell ATAC-seq data and available nascent transcription data, I was able to probe the extent of conserved enhancer characteristics on multiple levels. When compared with available single cell-type datasets for *Homo sapiens* and *Drosophila melanogaster*, we find that the traditional animal enhancer profile is not conserved in *Arabidopsis*. While many hyperacessible chromatin sites exist across the plant's genome, both within and around genes, they are not surrounded by well-positioned, modified nucleosomes as they are in animals. Transcription preferentially proceeds unidirectionally in *Arabidopsis*, leading to histone PTM deposition in the sense direction alone, both at gene bodies and at distal intergenic hyperaccessible sites. This trend makes it prohibitively challenging to identify putative enhancer elements in *Arabidopsis* by these criteria, indicating that the enhancer characteristics found to be conserved in animal species do not extend to all complex, multicellular eukaryotes. This suggests there is a need for innovative approaches to investigate transcriptional regulation in plants.

My dissertation has showcased a variety of approaches to identify and characterize putative enhancer elements in plant species. In the **Discussion** chapter of this dissertation, I expand upon the conclusions of our work and how it fits with the current state of the field. Notably, work published this

year examined the enrichment of a wide range of histone posttranslational modifications across several plant species (Lu, Marand et al. 2019, Ricci, Lu et al. 2019, Yan, Chen et al. 2019). While the results and implications of these groundbreaking studies will be explored further in the **Discussion** of this dissertation, they share the pitfall of earlier next-generation sequencing approaches in that they fail to take into account differences in enhancer activity state – and hence, associated histone modifications – due to differing cell types. Our approach utilizes the nuclei affinity purification technique of INTACT to examine histone PTM enrichment in a way that is high-throughput, time efficient, and – most importantly – is cell-type specific. Further in the **Discussion** I expand upon the future directions of our work. I present unpublished work we have generated in collaboration with the Queitsch lab to adapt Self-Transcribing Accessible Regulatory Region with high-throughput sequencing (STARR-seq) for use in plant species. STARR-seq is a high-throughput assay that can identify putative enhancers based on functionality, rather than secondary characteristics. Because our data, in combination with other studies recently published, point to significant differences between the qualities of animal and plant enhancers, it is doubtful that further investigations based on previous knowledge of animal regulatory elements will yield much success in the Plantae kingdom. Instead, STARR-seq represents an approach to identify thousands of putative enhancer sites simultaneously and in a relatively unbiased manner. Finally, I explore potential next steps for the advancement of enhancer research in plants.

FIGURES

Note: All figures in this section are original works and were created by the author, Kelsey A. Maher.

Figure 1.1

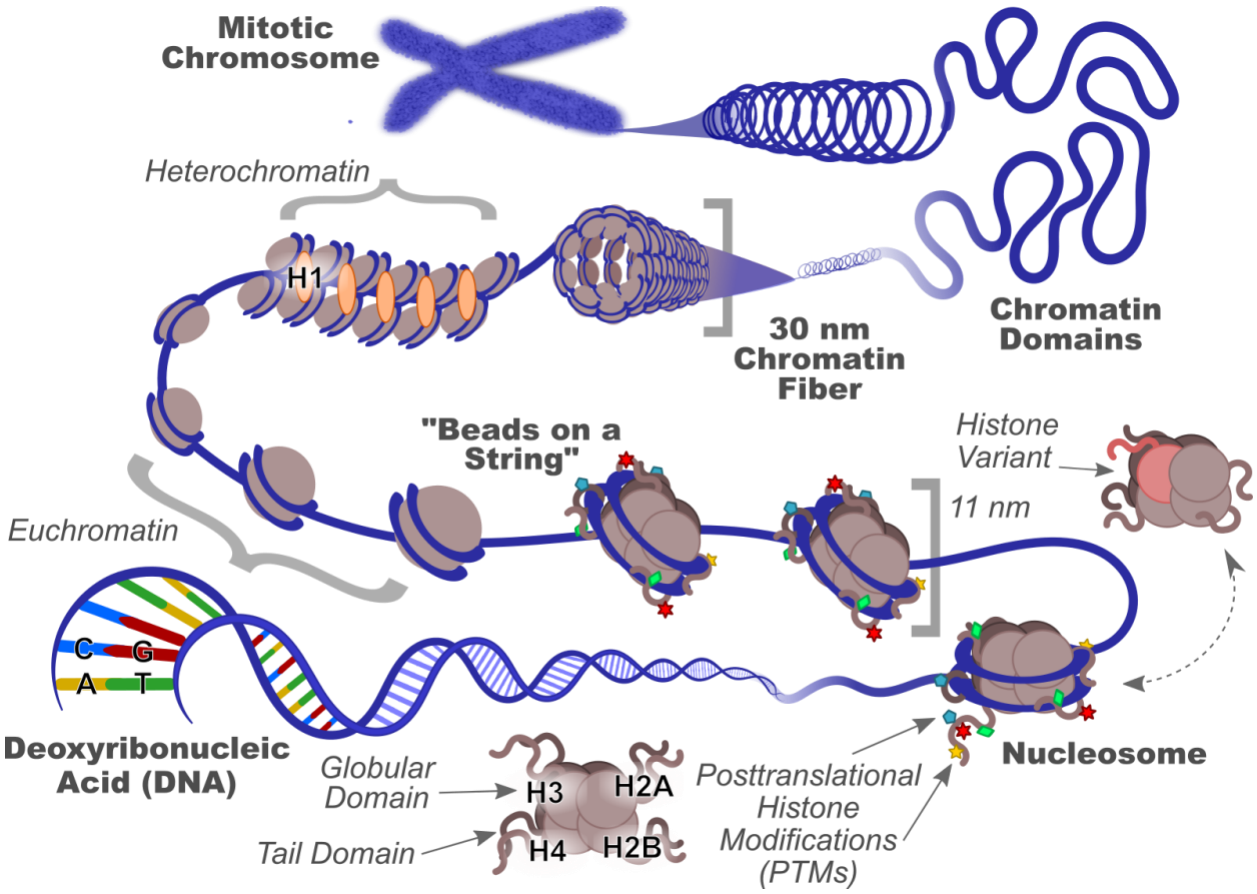


Figure 1.1. DNA Compaction within the eukaryotic cell. The genetic information within a cell's nucleus is highly compacted. The molecules of deoxyribonucleic acid (DNA), organized into a double-helix comprised of a sugar-phosphate backbone and nucleotide bases (adenosine (A), thymine (T), cytosine (C), and guanine (G)), are wrapped around octamers of histone proteins. This unit is called a nucleosome, and is comprised of ~147 bp of DNA, two H2A-H2B histone dimers, and a dimer of histone H3-H4 dimers, though variant histones may be substituted for particular cellular processes. Histone proteins are organized

into a central globular domain and a peripheral tail domain, both of which can be altered by chemical posttranslational modifications (PTMs) which impact the protein's ability to bind to DNA. The 11 nanometer (nm) array of nucleosomes is commonly referred to as "beads on a string" due to its characteristic appearance. Regions of nucleosomes that allow ready access to the underlying DNA, either via their generous spacing or by binding in an equilibrium with other *trans*-acting factors, are referred to as euchromatin. With the help of linker histone H1 and other architectural proteins, chromatin is condensed into higher compaction states, including heterochromatin, 30 nm chromatin fibers, chromatin domains, and ultimately mitotic chromosomes.

Figure 1.2

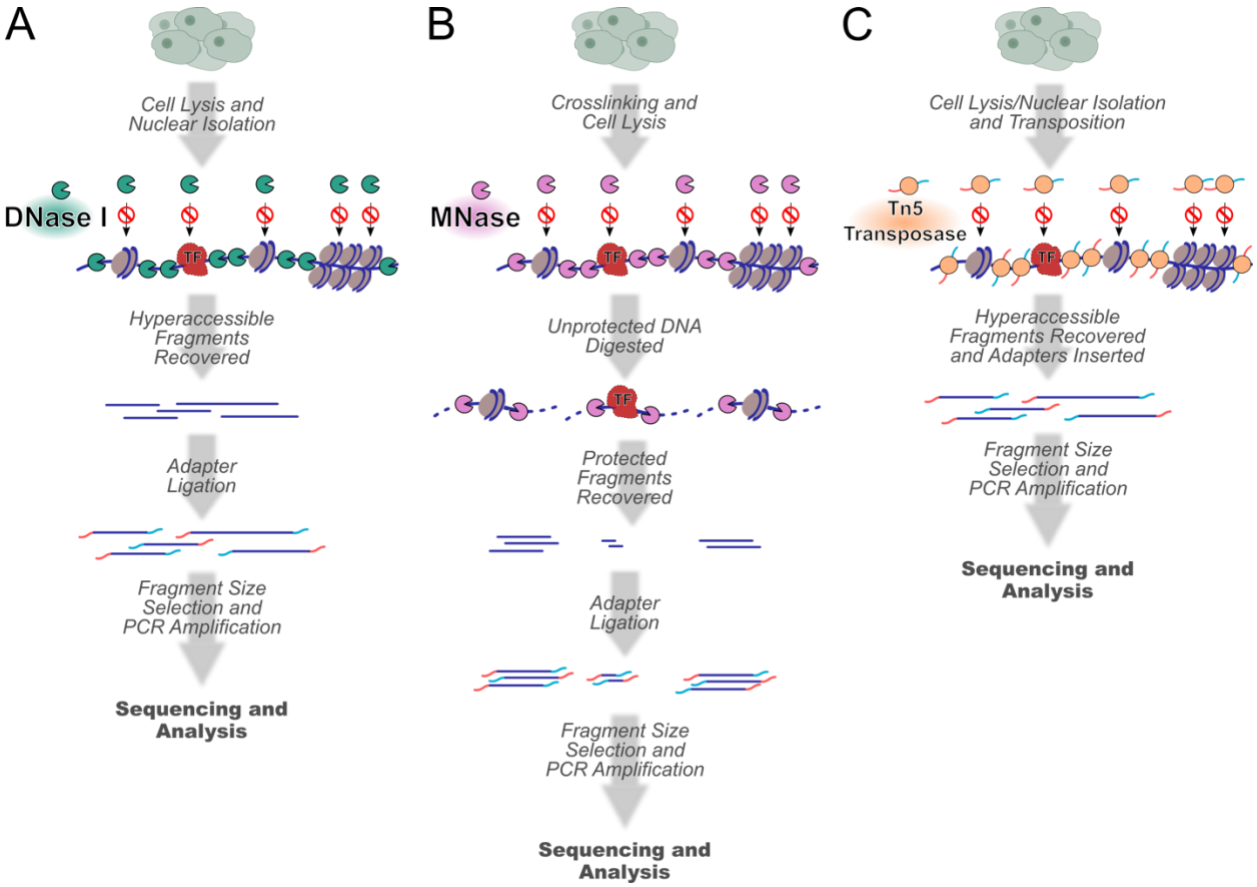


Figure 1.2. Comparison of the workflow of next-generation sequencing assays for chromatin accessibility. Outline of the steps involved to take organic material to sequenced dataset by **A)** DNase-seq, **B)** MNase-seq, and **C)** ATAC-seq. The key enzymes in these assays preferentially interact with accessible DNA – DNA not protected by bound nucleosomes, heterochromatin, or transcription factors (TFs). **A)** DNase-seq relies on the endonuclease DNase I, which cleaves the chromatin into fragments. Sequencing adapters are then ligated to these fragments, which then undergo size selection and PCR amplification before being submitted to deep sequencing. **B)** MNase-seq relies on the endo-exonuclease micrococcal nuclease which cleaves the chromatin and digests overhang fragments. This results in nucleosomal-sized fragments (~150 bp) or smaller fragments reflective of a transcription factor footprint. Similar to DNase-seq, these fragments are size-selected, amplified, and sequenced. **C)** ATAC-seq relies on a hyperactive

Tn5 transposase pre-loaded with sequencing adapters. The enzyme carries out a transposition reaction at the accessible chromatin site, simultaneous cleaving the DNA and attaching the sequencing adapters. Because fragmentation and adapter attachment take place in a single step, ATAC-seq libraries can be prepared for sequencing much more easily and quickly than DNase-seq or MNase-seq.

Figure 1.3

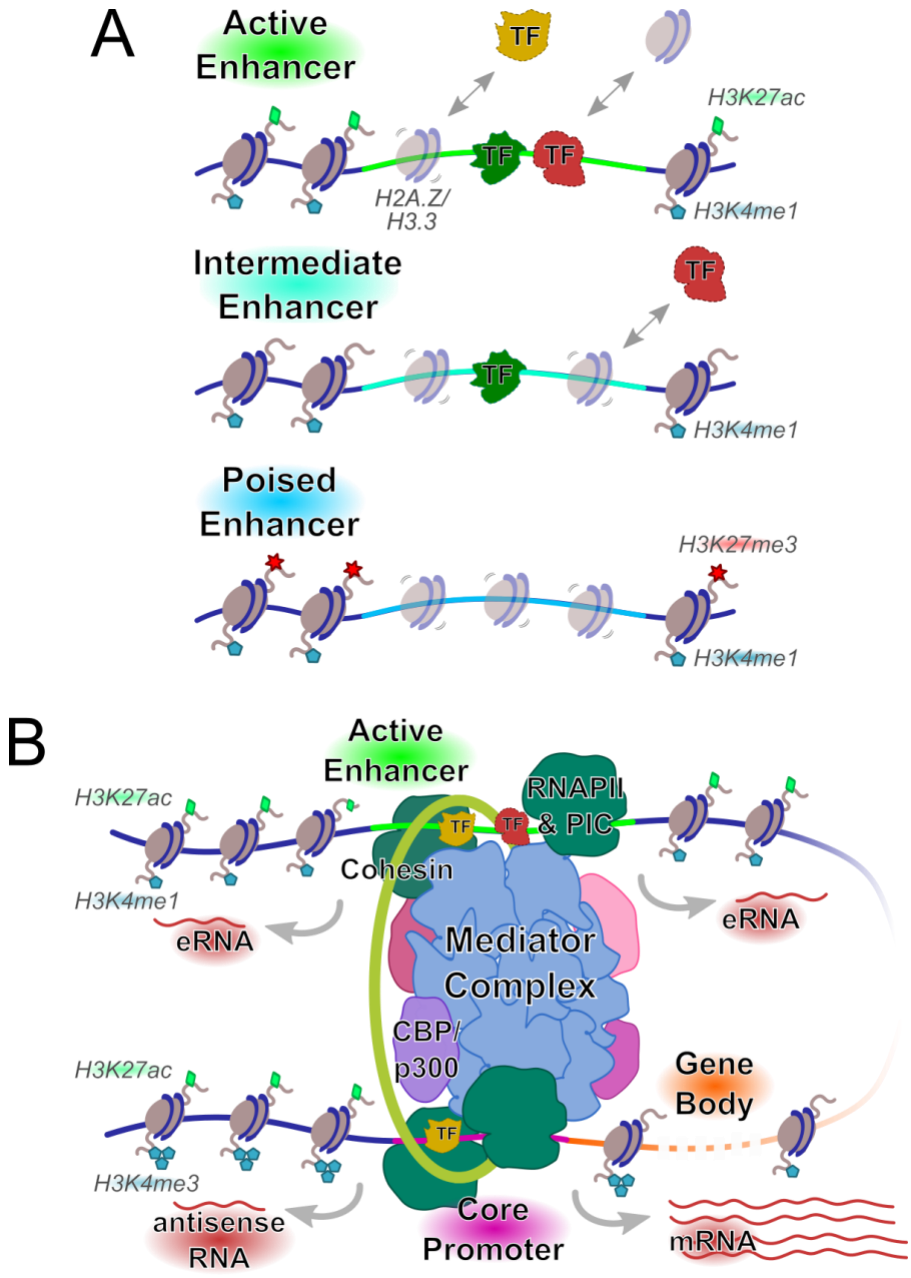


Figure 1.3. Enhancers have multiple states of activity and interact with target promoters to amplify transcription. A) Enhancers have been described to have multiple activity states, marked by different epigenetic modifications. 'Active' enhancers are flanked by well-positioned nucleosomes with H3K4me1 and H3K27ac, 'Intermediate' enhancers are flanked by nucleosomes with H3K4me1, and 'poised' enhancers are flanked with nucleosomes with H3K4me1 and H3K27me3. Enhancers reside in

'nucleosome-depleted regions' which are enriched with unstable H2A.Z/H3.2 nucleosomes. These nucleosomes are easily evicted, and exist in a state of equilibrium with DNA-binding transcription factors (TFs). **B)** An active enhancer will maneuver in space to associate with its target promoter. The interaction is stabilized by cohesion and the mediator complex, and recruits histone-modifying machinery such as histone acetylase CBP/p300. When activated, RNA polymerase II (RNAPII) in the assembled preinitiation complex (PIC) will initiate transcription, producing RNA products bidirectionally from both the promoter and the enhancer. The enhancer transcripts, called 'eRNA', are typically short and bidirectional, as are the antisense transcripts produced upstream of the promoter. mRNAs are long, unidirectional, and polyadenylated, and are produced at levels much higher than background due to the influence of the enhancer. The body of active genes is preferentially enriched for H3K4me3 towards the 5' end where transcriptional activity levels are the highest, and enriched for H3K4me1 towards the 3' end as transcriptional activity levels dwindle.

LITERATURE CITED

1. Luger K, Mader A, Sargent DF, Richmond TJ. The atomic structure of the nucleosome core particle. J Biomol Struct Dyn. 2000;17 Suppl 1:185-8. Epub 2000/01/01. doi: 10.1080/07391102.2000.10506619. PubMed PMID: 22607422.
2. Kornberg RD. Structure of chromatin. Annu Rev Biochem. 1977;46:931-54. Epub 1977/01/01. doi: 10.1146/annurev.bi.46.070177.004435. PubMed PMID: 332067.
3. Arents G, Burlingame RW, Wang BC, Love WE, Moudrianakis EN. The nucleosomal core histone octamer at 3.1 A resolution: a tripartite protein assembly and a left-handed superhelix. Proc Natl Acad Sci U S A. 1991;88(22):10148-52. Epub 1991/11/15. doi: 10.1073/pnas.88.22.10148. PubMed PMID: 1946434; PubMed Central PMCID: PMCPMC52885.
4. Marzluff WF, Duronio RJ. Histone mRNA expression: multiple levels of cell cycle regulation and important developmental consequences. Curr Opin Cell Biol. 2002;14(6):692-9. Epub 2002/12/11. doi: 10.1016/s0955-0674(02)00387-3. PubMed PMID: 12473341.
5. Marzluff WF, Gongidi P, Woods KR, Jin J, Maltais LJ. The human and mouse replication-dependent histone genes. Genomics. 2002;80(5):487-98. Epub 2002/11/01. PubMed PMID: 12408966.
6. Clapier CR, Cairns BR. The biology of chromatin remodeling complexes. Annu Rev Biochem. 2009;78:273-304. Epub 2009/04/10. doi: 10.1146/annurev.biochem.77.062706.153223. PubMed PMID: 19355820.
7. Palmer DK, O'Day K, Trong HL, Charbonneau H, Margolis RL. Purification of the centromere-specific protein CENP-A and demonstration that it is a distinctive histone. Proc Natl Acad Sci U S A. 1991;88(9):3734-8. Epub 1991/05/01. doi: 10.1073/pnas.88.9.3734. PubMed PMID: 2023923; PubMed Central PMCID: PMCPMC51527.
8. Morrison AJ, Shen X. Chromatin remodelling beyond transcription: the INO80 and SWR1 complexes. Nat Rev Mol Cell Biol. 2009;10(6):373-84. Epub 2009/05/09. doi: 10.1038/nrm2693. PubMed PMID: 19424290; PubMed Central PMCID: PMCPMC6103619.

9. Lowndes NF, Toh GW. DNA repair: the importance of phosphorylating histone H2AX. Curr Biol. 2005;15(3):R99-r102. Epub 2005/02/08. doi: 10.1016/j.cub.2005.01.029. PubMed PMID: 15694301.

10. Olins AL, Olins DE. Spheroid chromatin units (v bodies). Science. 1974;183(4122):330-2. Epub 1974/01/25. doi: 10.1126/science.183.4122.330. PubMed PMID: 4128918.

11. Luger K, Dechassa ML, Tremethick DJ. New insights into nucleosome and chromatin structure: an ordered state or a disordered affair? Nat Rev Mol Cell Biol. 2012;13(7):436-47. Epub 2012/06/23. doi: 10.1038/nrm3382. PubMed PMID: 22722606; PubMed Central PMCID: PMCPMC3408961.

12. Robinson PJ, Rhodes D. Structure of the '30 nm' chromatin fibre: a key role for the linker histone. Curr Opin Struct Biol. 2006;16(3):336-43. Epub 2006/05/23. doi: 10.1016/j.sbi.2006.05.007. PubMed PMID: 16714106.

13. Canzio D, Chang EY, Shankar S, Kuchenbecker KM, Simon MD, Madhani HD, Narlikar GJ, Al-Sady B. Chromodomain-mediated oligomerization of HP1 suggests a nucleosome-bridging mechanism for heterochromatin assembly. Mol Cell. 2011;41(1):67-81. Epub 2011/01/08. doi: 10.1016/j.molcel.2010.12.016. PubMed PMID: 21211724; PubMed Central PMCID: PMCPMC3752404.

14. Francis NJ, Kingston RE, Woodcock CL. Chromatin compaction by a polycomb group protein complex. Science. 2004;306(5701):1574-7. Epub 2004/11/30. doi: 10.1126/science.1100576. PubMed PMID: 15567868.

15. Song F, Chen P, Sun D, Wang M, Dong L, Liang D, Xu RM, Zhu P, Li G. Cryo-EM study of the chromatin fiber reveals a double helix twisted by tetranucleosomal units. Science. 2014;344(6182):376-80. Epub 2014/04/26. doi: 10.1126/science.1251413. PubMed PMID: 24763583.

16. Amendola M, van Steensel B. Mechanisms and dynamics of nuclear lamina-genome interactions. Curr Opin Cell Biol. 2014;28:61-8. Epub 2014/04/04. doi: 10.1016/j.ceb.2014.03.003. PubMed PMID: 24694724.

17. Gonzalez-Sandoval A, Gasser SM. On TADs and LADs: Spatial Control Over Gene Expression. Trends Genet. 2016;32(8):485-95. Epub 2016/06/18. doi: 10.1016/j.tig.2016.05.004. PubMed PMID: 27312344.

18. Hirano T. Condensins: universal organizers of chromosomes with diverse functions. Genes Dev. 2012;26(15):1659-78. Epub 2012/08/03. doi: 10.1101/gad.194746.112. PubMed PMID: 22855829; PubMed Central PMCID: PMCPMC3418584.

19. Nasmyth K, Haering CH. Cohesin: its roles and mechanisms. Annu Rev Genet. 2009;43:525-58. Epub 2009/11/06. doi: 10.1146/annurev-genet-102108-134233. PubMed PMID: 19886810.

20. Allshire RC, Ekwall K. Epigenetic Regulation of Chromatin States in Schizosaccharomyces pombe. Cold Spring Harb Perspect Biol. 2015;7(7):a018770. Epub 2015/07/03. doi: 10.1101/cshperspect.a018770. PubMed PMID: 26134317; PubMed Central PMCID: PMCPMC4484966.

21. Martienssen R, Moazed D. RNAi and heterochromatin assembly. Cold Spring Harb Perspect Biol. 2015;7(8):a019323. Epub 2015/08/05. doi: 10.1101/cshperspect.a019323. PubMed PMID: 26238358; PubMed Central PMCID: PMCPMC4526745.

22. Pikaard CS, Mittelsten Scheid O. Epigenetic regulation in plants. Cold Spring Harb Perspect Biol. 2014;6(12):a019315. Epub 2014/12/03. doi: 10.1101/cshperspect.a019315. PubMed PMID: 25452385; PubMed Central PMCID: PMCPMC4292151.

23. Strome S, Kelly WG, Ercan S, Lieb JD. Regulation of the X chromosomes in Caenorhabditis elegans. Cold Spring Harb Perspect Biol. 2014;6(3). Epub 2014/03/05. doi: 10.1101/cshperspect.a018366. PubMed PMID: 24591522; PubMed Central PMCID: PMCPMC3942922.

24. Sherwood RI, Hashimoto T, O'Donnell CW, Lewis S, Barkal AA, van Hoff JP, Karun V, Jaakkola T, Gifford DK. Discovery of directional and nondirectional pioneer transcription factors by modeling DNase profile magnitude and shape. Nat Biotechnol. 2014;32(2):171-8. Epub 2014/01/21. doi: 10.1038/nbt.2798. PubMed PMID: 24441470; PubMed Central PMCID: PMCPMC3951735.

25. Soufi A, Garcia MF, Jaroszewicz A, Osman N, Pellegrini M, Zaret KS. Pioneer transcription factors target partial DNA motifs on nucleosomes to initiate reprogramming. Cell. 2015;161(3):555-68. Epub 2015/04/22. doi: 10.1016/j.cell.2015.03.017. PubMed PMID: 25892221; PubMed Central PMCID: PMCPMC4409934.

26. Henikoff S, Henikoff JG, Sakai A, Loeb GB, Ahmad K. Genome-wide profiling of salt fractions maps physical properties of chromatin. Genome Res. 2009;19(3):460-9. Epub 2008/12/18. doi: 10.1101/gr.087619.108. PubMed PMID: 19088306; PubMed Central PMCID: PMCPMC2661814.

27. Jin C, Zang C, Wei G, Cui K, Peng W, Zhao K, Felsenfeld G. H3.3/H2A.Z double variant-containing nucleosomes mark 'nucleosome-free regions' of active promoters and other regulatory regions. Nat Genet. 2009;41(8):941-5. Epub 2009/07/28. doi: 10.1038/ng.409. PubMed PMID: 19633671; PubMed Central PMCID: PMCPMC3125718.

28. Jin C, Felsenfeld G. Nucleosome stability mediated by histone variants H3.3 and H2A.Z. Genes Dev. 2007;21(12):1519-29. Epub 2007/06/19. doi: 10.1101/gad.1547707. PubMed PMID: 17575053; PubMed Central PMCID: PMCPMC1891429.

29. Boyle AP, Davis S, Shulha HP, Meltzer P, Margulies EH, Weng Z, Furey TS, Crawford GE. High-resolution mapping and characterization of open chromatin across the genome. Cell. 2008;132(2):311-22. Epub 2008/02/05. doi: 10.1016/j.cell.2007.12.014. PubMed PMID: 18243105; PubMed Central PMCID: PMCPMC2669738.

30. Yuan GC, Liu YJ, Dion MF, Slack MD, Wu LF, Altschuler SJ, Rando OJ. Genome-scale identification of nucleosome positions in S. cerevisiae. Science. 2005;309(5734):626-30. Epub 2005/06/18. doi: 10.1126/science.1112178. PubMed PMID: 15961632.

31. Buenrostro JD, Giresi PG, Zaba LC, Chang HY, Greenleaf WJ. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. Nature methods. 2013;10(12):1213-8. Epub 2013/10/08. doi: 10.1038/nmeth.2688. PubMed PMID: 24097267; PubMed Central PMCID: PMCPmc3959825.

32. Buenrostro JD, Wu B, Chang HY, Greenleaf WJ. ATAC-seq: A Method for Assaying Chromatin Accessibility Genome-Wide. Curr Protoc Mol Biol. 2015;109:21.9.1-9. Epub 2015/01/07. doi: 10.1002/0471142727.mb2129s109. PubMed PMID: 25559105; PubMed Central PMCID: PMCPmc4374986.

33. Bernstein BE, Mikkelsen TS, Xie X, Kamal M, Huebert DJ, Cuff J, Fry B, Meissner A, Wernig M, Plath K, Jaenisch R, Wagschal A, Feil R, Schreiber SL, Lander ES. A bivalent chromatin structure marks key developmental genes in embryonic stem cells. Cell. 2006;125(2):315-26. Epub 2006/04/25. doi: 10.1016/j.cell.2006.02.041. PubMed PMID: 16630819.

34. Sims RJ, 3rd, Belotserkovskaya R, Reinberg D. Elongation by RNA polymerase II: the short and long of it. Genes Dev. 2004;18(20):2437-68. Epub 2004/10/19. doi: 10.1101/gad.1235904. PubMed PMID: 15489290.

35. Voss TC, Hager GL. Dynamic regulation of transcriptional states by chromatin and transcription factors. Nature reviews Genetics. 2014;15(2):69-81. Epub 2013/12/18. doi: 10.1038/nrg3623. PubMed PMID: 24342920; PubMed Central PMCID: PMCPMC6322398.

36. Spitz F, Furlong EE. Transcription factors: from enhancer binding to developmental control. Nature reviews Genetics. 2012;13(9):613-26. Epub 2012/08/08. doi: 10.1038/nrg3207. PubMed PMID: 22868264.

37. Ayer S, Benyajati C. Conserved enhancer and silencer elements responsible for differential Adh transcription in Drosophila cell lines. Mol Cell Biol. 1990;10(7):3512-23. Epub 1990/07/01. doi: 10.1128/mcb.10.7.3512. PubMed PMID: 1694013; PubMed Central PMCID: PMCPMC360786.

38. Petrykowska HM, Vockley CM, Elnitski L. Detection and characterization of silencers and enhancer-blockers in the greater CFTR locus. Genome Res. 2008;18(8):1238-46. Epub 2008/04/26. doi: 10.1101/gr.073817.107. PubMed PMID: 18436892; PubMed Central PMCID: PMCPMC2493434.

39. Vokes SA, Ji H, Wong WH, McMahon AP. A genome-scale analysis of the cis-regulatory circuitry underlying sonic hedgehog-mediated patterning of the mammalian limb. Genes Dev. 2008;22(19):2651-

63. Epub 2008/10/04. doi: 10.1101/gad.1693008. PubMed PMID: 18832070; PubMed Central PMCID: PMCPMC2559910.

40. Gaszner M, Felsenfeld G. Insulators: exploiting transcriptional and epigenetic mechanisms. Nature reviews Genetics. 2006;7(9):703-13. Epub 2006/08/16. doi: 10.1038/nrg1925. PubMed PMID: 16909129.

41. Schwaiger M, Schonauer A, Rendeiro AF, Pribitzer C, Schauer A, Gilles AF, Schinko JB, Renfer E, Fredman D, Technau U. Evolutionary conservation of the eumetazoan gene regulatory landscape. Genome Res. 2014;24(4):639-50. Epub 2014/03/20. doi: 10.1101/gr.162529.113. PubMed PMID: 24642862; PubMed Central PMCID: PMCPmc3975063.

42. Villar D, Berthelot C, Aldridge S, Rayner TF, Lukk M, Pignatelli M, Park TJ, Deaville R, Erichsen JT, Jasinska AJ, Turner JM, Bertelsen MF, Murchison EP, Flicek P, Odom DT. Enhancer evolution across 20 mammalian species. Cell. 2015;160(3):554-66. Epub 2015/01/31. doi: 10.1016/j.cell.2015.01.006. PubMed PMID: 25635462; PubMed Central PMCID: PMCPmc4313353.

43. Zhu B, Zhang W, Zhang T. Genome-Wide Prediction and Validation of Intergenic Enhancers in Arabidopsis Using Open Chromatin Signatures2015;27(9):2415-26. doi: 10.1105/tpc.15.00537. PubMed PMID: 26373455.

44. Weber B, Zicola J, Oka R, Stam M. Plant Enhancers: A Call for Discovery. Trends in plant science. 2016;21(11):974-87. Epub 2016/10/31. doi: 10.1016/j.tplants.2016.07.013. PubMed PMID: 27593567.

45. Xu H, Hoover TR. Transcriptional regulation at a distance in bacteria. Curr Opin Microbiol. 2001;4(2):138-44. Epub 2001/04/03. PubMed PMID: 11282468.

46. Berg PE, Popovic Z, Anderson WF. Promoter dependence of enhancer activity. Mol Cell Biol. 1984;4(8):1664-8. Epub 1984/08/01. PubMed PMID: 6092930; PubMed Central PMCID: PMCPmc368966.

47. Lee TI, Young RA. Transcription of eukaryotic protein-coding genes. Annu Rev Genet. 2000;34:77-137. Epub 2000/11/28. doi: 10.1146/annurev.genet.34.1.77. PubMed PMID: 11092823.

48. Sainsbury S, Bernecky C, Cramer P. Structural basis of transcription initiation by RNA polymerase II. Nat Rev Mol Cell Biol. 2015;16(3):129-43. Epub 2015/02/19. doi: 10.1038/nrm3952. PubMed PMID: 25693126.

49. Kagey MH, Newman JJ, Bilodeau S, Zhan Y, Orlando DA, van Berkum NL, Ebmeier CC, Goossens J, Rahl PB, Levine SS, Taatjes DJ, Dekker J, Young RA. Mediator and cohesin connect gene expression and chromatin architecture. Nature. 2010;467(7314):430-5. Epub 2010/08/20. doi: 10.1038/nature09380. PubMed PMID: 20720539; PubMed Central PMCID: PMCPMC2953795.

50. Ebmeier CC, Taatjes DJ. Activator-Mediator binding regulates Mediator-cofactor interactions. Proc Natl Acad Sci U S A. 2010;107(25):11283-8. Epub 2010/06/11. doi: 10.1073/pnas.0914215107. PubMed PMID: 20534441; PubMed Central PMCID: PMCPMC2895140.

51. Poss ZC, Ebmeier CC, Taatjes DJ. The Mediator complex and transcription regulation. Crit Rev Biochem Mol Biol. 2013;48(6):575-608. Epub 2013/10/04. doi: 10.3109/10409238.2013.840259. PubMed PMID: 24088064; PubMed Central PMCID: PMCPMC3852498.

52. Schmidt D, Schwalie PC, Ross-Innes CS, Hurtado A, Brown GD, Carroll JS, Flicek P, Odom DT. A CTCF-independent role for cohesin in tissue-specific transcription. Genome Res. 2010;20(5):578-88. Epub 2010/03/12. doi: 10.1101/gr.100479.109. PubMed PMID: 20219941; PubMed Central PMCID: PMCPMC2860160.

53. Faure AJ, Schmidt D, Watt S, Schwalie PC, Wilson MD, Xu H, Ramsay RG, Odom DT, Flicek P. Cohesin regulates tissue-specific expression by stabilizing highly occupied cis-regulatory modules. Genome Res. 2012;22(11):2163-75. Epub 2012/07/12. doi: 10.1101/gr.136507.111. PubMed PMID: 22780989; PubMed Central PMCID: PMCPMC3483546.

54. Vernimmen D, Bickmore WA. The Hierarchy of Transcriptional Activation: From Enhancer to Promoter. Trends Genet. 2015;31(12):696-708. Epub 2015/11/26. doi: 10.1016/j.tig.2015.10.004. PubMed PMID: 26599498.

55. Ghavi-Helm Y, Klein FA, Pakozdi T, Ciglar L, Noordermeer D, Huber W, Furlong EE. Enhancer loops appear stable during development and are associated with paused polymerase. Nature. 2014;512(7512):96-100. Epub 2014/07/22. doi: 10.1038/nature13417. PubMed PMID: 25043061.

56. Tolhuis B, Palstra RJ, Splinter E, Grosveld F, de Laat W. Looping and interaction between hypersensitive sites in the active beta-globin locus. Mol Cell. 2002;10(6):1453-65. Epub 2002/12/31. doi: 10.1016/s1097-2765(02)00781-5. PubMed PMID: 12504019.

57. Carter D, Chakalova L, Osborne CS, Dai YF, Fraser P. Long-range chromatin regulatory interactions in vivo. Nat Genet. 2002;32(4):623-6. Epub 2002/11/12. doi: 10.1038/ng1051. PubMed PMID: 12426570.

58. Kulaeva OI, Nizovtseva EV, Polikanov YS, Ulianov SV, Studitsky VM. Distant activation of transcription: mechanisms of enhancer action. Mol Cell Biol. 2012;32(24):4892-7. Epub 2012/10/10. doi: 10.1128/mcb.01127-12. PubMed PMID: 23045397; PubMed Central PMCID: PMCPMC3510544.

59. Lemon B, Tjian R. Orchestrated response: a symphony of transcription factors for gene control. Genes Dev. 2000;14(20):2551-69. Epub 2000/10/21. doi: 10.1101/gad.831000. PubMed PMID: 11040209.

60. Kim TK, Shiekhattar R. Architectural and Functional Commonalities between Enhancers and Promoters. Cell. 2015;162(5):948-59. Epub 2015/09/01. doi: 10.1016/j.cell.2015.08.008. PubMed PMID: 26317464; PubMed Central PMCID: PMCPMC4556168.

61. Anders S, Huber W. Differential expression analysis for sequence count data. Genome Biol. 2010;11(10):R106. Epub 2010/10/29. doi: 10.1186/gb-2010-11-10-r106. PubMed PMID: 20979621; PubMed Central PMCID: PMCPMC3218662.

62. Creyghton MP, Cheng AW, Welstead GG, Kooistra T, Carey BW, Steine EJ, Hanna J, Lodato MA, Frampton GM, Sharp PA, Boyer LA, Young RA, Jaenisch R. Histone H3K27ac separates active from poised enhancers and predicts developmental state. Proc Natl Acad Sci U S A. 2010;107(50):21931-6. Epub 2010/11/26. doi: 10.1073/pnas.1016071107. PubMed PMID: 21106759; PubMed Central PMCID: PMCPmc3003124.

63. Heintzman ND, Hon GC, Hawkins RD, Kheradpour P, Stark A, Harp LF, Ye Z, Lee LK, Stuart RK, Ching CW, Ching KA, Antosiewicz-Bourget JE, Liu H, Zhang X, Green RD, Lobanenkov VV, Stewart R, Thomson JA, Crawford GE, Kellis M, Ren B. Histone modifications at human enhancers reflect global cell-type-specific gene expression. Nature. 2009;459(7243):108-12. Epub 2009/03/20. doi: 10.1038/nature07829. PubMed PMID: 19295514; PubMed Central PMCID: PMCPmc2910248.

64. Kleinjan DA, van Heyningen V. Long-range control of gene expression: emerging mechanisms and disruption in disease. Am J Hum Genet. 2005;76(1):8-32. Epub 2004/11/19. doi: 10.1086/426833. PubMed PMID: 15549674; PubMed Central PMCID: PMCPmc1196435.

65. Lettice LA, Heaney SJ, Purdie LA, Li L, de Beer P, Oostra BA, Goode D, Elgar G, Hill RE, de Graaff E. A long-range Shh enhancer regulates expression in the developing limb and fin and is associated with preaxial polydactyly. Hum Mol Genet. 2003;12(14):1725-35. Epub 2003/07/03. PubMed PMID: 12837695.

66. Ong CT, Corces VG. Enhancer function: new insights into the regulation of tissue-specific gene expression. Nature reviews Genetics. 2011;12(4):283-93. Epub 2011/03/02. doi: 10.1038/nrg2957. PubMed PMID: 21358745; PubMed Central PMCID: PMCPmc3175006.

67. The ENCODE (ENCyclopedia Of DNA Elements) Project. Science. 2004;306(5696):636-40. Epub 2004/10/23. doi: 10.1126/science.1105136. PubMed PMID: 15499007.

68. An integrated encyclopedia of DNA elements in the human genome. Nature. 2012;489(7414):57-74. Epub 2012/09/08. doi: 10.1038/nature11247. PubMed PMID: 22955616; PubMed Central PMCID: PMCPMC3439153.

69. Boley N, Wan KH, Bickel PJ, Celniker SE. Navigating and mining modENCODE data. Methods. 2014;68(1):38-47. Epub 2014/03/19. doi: 10.1016/j.ymeth.2014.03.007. PubMed PMID: 24636835; PubMed Central PMCID: PMCPMC4857704.

70. Birney E, Stamatoyannopoulos JA, Dutta A, Guigó R, Gingeras TR, Margulies EH, Weng Z, Snyder M, Dermitzakis ET, Thurman RE, Kuehn MS, Taylor CM, Neph S, Koch CM, Asthana S, Malhotra A, Adzhubei I, Greenbaum JA, Andrews RM, Flicek P, Boyle PJ, Cao H, Carter NP, Clelland

GK, Davis S, Day N, Dhami P, Dillon SC, Dorschner MO, Fiegler H, Giresi PG, Goldy J, Hawrylycz M, Haydock A, Humbert R, James KD, Johnson BE, Johnson EM, Frum TT, Rosenzweig ER, Karnani N, Lee K, Lefebvre GC, Navas PA, Neri F, Parker SC, Sabo PJ, Sandstrom R, Shafer A, Vetrie D, Weaver M, Wilcox S, Yu M, Collins FS, Dekker J, Lieb JD, Tullius TD, Crawford GE, Sunyaev S, Noble WS, Dunham I, Denoeud F, Reymond A, Kapranov P, Rozowsky J, Zheng D, Castelo R, Frankish A, Harrow J, Ghosh S, Sandelin A, Hofacker IL, Baertsch R, Keefe D, Dike S, Cheng J, Hirsch HA, Sekinger EA, Lagarde J, Abril JF, Shahab A, Flamm C, Fried C, Hackermüller J, Hertel J, Lindemeyer M, Missal K, Tanzer A, Washietl S, Korbel J, Emanuelsson O, Pedersen JS, Holroyd N, Taylor R, Swarbreck D, Matthews N, Dickson MC, Thomas DJ, Weirauch MT, Gilbert J, Drenkow J, Bell I, Zhao X, Srinivasan KG, Sung WK, Ooi HS, Chiu KP, Foissac S, Alioto T, Brent M, Pachter L, Tress ML, Valencia A, Choo SW, Choo CY, Ucla C, Manzano C, Wyss C, Cheung E, Clark TG, Brown JB, Ganesh M, Patel S, Tammana H, Chrast J, Henrichsen CN, Kai C, Kawai J, Nagalakshmi U, Wu J, Lian Z, Lian J, Newburger P, Zhang X, Bickel P, Mattick JS, Carninci P, Hayashizaki Y, Weissman S, Hubbard T, Myers RM, Rogers J, Stadler PF, Lowe TM, Wei CL, Ruan Y, Struhl K, Gerstein M, Antonarakis SE, Fu Y, Green ED, Karaöz U, Siepel A, Taylor J, Liefer LA, Wetterstrand KA, Good PJ, Feingold EA, Guyer MS, Cooper GM, Asimenos G, Dewey CN, Hou M, Nikolaev S, Montoya-Burgos JI, Löytynoja A, Whelan S, Pardi F, Massingham T, Huang H, Zhang NR, Holmes I, Mullikin JC, Ureta-Vidal A, Paten B, Seringhaus M, Church D, Rosenbloom K, Kent WJ, Stone EA, Batzoglou S, Goldman N, Hardison RC, Haussler D, Miller W, Sidow A, Trinklein ND, Zhang ZD, Barrera L, Stuart R, King DC, Ameur A, Enroth S, Bieda MC, Kim J, Bhinge AA, Jiang N, Liu J, Yao F, Vega VB, Lee CW, Ng P, Shahab A, Yang A, Moqtaderi Z, Zhu Z, Xu X, Squazzo S, Oberley MJ, Inman D, Singer MA, Richmond TA, Munn KJ, Rada-Iglesias A, Wallerman O, Komorowski J, Fowler JC, Couttet P, Bruce AW, Dovey OM, Ellis PD, Langford CF, Nix DA, Euskirchen G, Hartman S, Urban AE, Kraus P, Van Calcar S, Heintzman N, Kim TH, Wang K, Qu C, Hon G, Luna R, Glass CK, Rosenfeld MG, Aldred SF, Cooper SJ, Halees A, Lin JM, Shulha HP, Zhang X, Xu M, Haidar JN, Yu Y, Ruan Y, Iyer VR, Green RD, Wadelius C, Farnham PJ, Ren B, Harte RA, Hinrichs AS, Trumbower H, Clawson H, Hillman-Jackson

J, Zweig AS, Smith K, Thakkapallayil A, Barber G, Kuhn RM, Karolchik D, Armengol L, Bird CP, de Bakker PI, Kern AD, Lopez-Bigas N, Martin JD, Stranger BE, Woodroffe A, Davydov E, Dimas A, Eyras E, Hallgrímsdóttir IB, Huppert J, Zody MC, Abecasis GR, Estivill X, Bouffard GG, Guan X, Hansen NF, Idol JR, Maduro VV, Maskeri B, McDowell JC, Park M, Thomas PJ, Young AC, Blakesley RW, Muzny DM, Sodergren E, Wheeler DA, Worley KC, Jiang H, Weinstock GM, Gibbs RA, Graves T, Fulton R, Mardis ER, Wilson RK, Clamp M, Cuff J, Gnerre S, Jaffe DB, Chang JL, Lindblad-Toh K, Lander ES, Koriabine M, Nefedov M, Osoegawa K, Yoshinaga Y, Zhu B, de Jong PJ. Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. Nature. 2007;447(7146):799-816. Epub 2007/06/16. doi: 10.1038/nature05874. PubMed PMID: 17571346; PubMed Central PMCID: PMCPMC2212820.

71. Davis CA, Hitz BC, Sloan CA, Chan ET, Davidson JM, Gabdank I, Hilton JA, Jain K, Baymuradov UK, Narayanan AK, Onate KC, Graham K, Miyasato SR, Dreszer TR, Strattan JS, Jolanki O, Tanaka FY, Cherry JM. The Encyclopedia of DNA elements (ENCODE): data portal update. Nucleic Acids Res. 2018;46(D1):D794-d801. Epub 2017/11/11. doi: 10.1093/nar/gkx1081. PubMed PMID: 29126249; PubMed Central PMCID: PMCPMC5753278.

72. Keene MA, Corces V, Lowenhaupt K, Elgin SC. DNase I hypersensitive sites in Drosophila chromatin occur at the 5' ends of regions of transcription. Proc Natl Acad Sci U S A. 1981;78(1):143-6. Epub 1981/01/01. PubMed PMID: 6264428; PubMed Central PMCID: PMCPMC319007.

73. McGhee JD, Wood WI, Dolan M, Engel JD, Felsenfeld G. A 200 base pair region at the 5' end of the chicken adult beta-globin gene is accessible to nuclease digestion. Cell. 1981;27(1 Pt 2):45-55. Epub 1981/11/01. PubMed PMID: 6276024.

74. Jiang J. The 'dark matter' in the plant genomes: non-coding and unannotated DNA sequences associated with open chromatin. Curr Opin Plant Biol. 2015;24:17-23. Epub 2015/01/28. doi: 10.1016/j.pbi.2015.01.005. PubMed PMID: 25625239.

75. Tsompana M, Buck MJ. Chromatin accessibility: a window into the genome. Epigenetics Chromatin. 2014;7(1):33. Epub 2014/12/05. doi: 10.1186/1756-8935-7-33. PubMed PMID: 25473421; PubMed Central PMCID: PMCPMC4253006.

76. Zentner GE, Scacheri PC. The chromatin fingerprint of gene enhancer elements. J Biol Chem. 2012;287(37):30888-96. Epub 2012/09/07. doi: 10.1074/jbc.R111.296491. PubMed PMID: 22952241; PubMed Central PMCID: PMCPMC3438921.

77. Heinz S, Romanoski CE, Benner C, Glass CK. The selection and function of cell type-specific enhancers. Nat Rev Mol Cell Biol. 2015;16(3):144-54. Epub 2015/02/05. doi: 10.1038/nrm3949. PubMed PMID: 25650801; PubMed Central PMCID: PMCPMC4517609.

78. Angeloni A, Bogdanovic O. Enhancer DNA methylation: implications for gene regulation. Essays Biochem. 2019;63(6):707-15. Epub 2019/09/26. doi: 10.1042/ebc20190030. PubMed PMID: 31551326.

79. Lai F, Shiekhattar R. Enhancer RNAs: the new molecules of transcription. Curr Opin Genet Dev. 2014;25:38-42. Epub 2014/02/01. doi: 10.1016/j.gde.2013.11.017. PubMed PMID: 24480293; PubMed Central PMCID: PMCPMC4484728.

80. Barski A, Cuddapah S, Cui K, Roh TY, Schones DE, Wang Z, Wei G, Chepelev I, Zhao K. High-resolution profiling of histone methylations in the human genome. Cell. 2007;129(4):823-37. Epub 2007/05/22. doi: 10.1016/j.cell.2007.05.009. PubMed PMID: 17512414.

81. Wang Z, Zang C, Rosenfeld JA, Schones DE, Barski A, Cuddapah S, Cui K, Roh TY, Peng W, Zhang MQ, Zhao K. Combinatorial patterns of histone acetylations and methylations in the human genome. Nat Genet. 2008;40(7):897-903. Epub 2008/06/17. doi: 10.1038/ng.154. PubMed PMID: 18552846; PubMed Central PMCID: PMCPMC2769248.

82. Bonn S, Zinzen RP, Girardot C, Gustafson EH, Perez-Gonzalez A, Delhomme N, Ghavi-Helm Y, Wilczynski B, Riddell A, Furlong EE. Tissue-specific analysis of chromatin state identifies temporal signatures of enhancer activity during embryonic development. Nat Genet. 2012;44(2):148-56. Epub 2012/01/11. doi: 10.1038/ng.1064. PubMed PMID: 22231485.

83. Ernst J, Kheradpour P, Mikkelsen TS, Shoresh N, Ward LD, Epstein CB, Zhang X, Wang L, Issner R, Coyne M, Ku M, Durham T, Kellis M, Bernstein BE. Mapping and analysis of chromatin state dynamics in nine human cell types. Nature. 2011;473(7345):43-9. Epub 2011/03/29. doi: 10.1038/nature09906. PubMed PMID: 21441907; PubMed Central PMCID: PMCPMC3088773.

84. Hawkins RD, Hon GC, Lee LK, Ngo Q, Lister R, Pelizzola M, Edsall LE, Kuan S, Luu Y, Klugman S, Antosiewicz-Bourget J, Ye Z, Espinoza C, Agarwahl S, Shen L, Ruotti V, Wang W, Stewart R, Thomson JA, Ecker JR, Ren B. Distinct epigenomic landscapes of pluripotent and lineage-committed human cells. Cell stem cell. 2010;6(5):479-91. Epub 2010/05/11. doi: 10.1016/j.stem.2010.03.018. PubMed PMID: 20452322; PubMed Central PMCID: PMCPMC2867844.

85. Zentner GE, Tesar PJ, Scacheri PC. Epigenetic signatures distinguish multiple classes of enhancers with distinct cellular functions. Genome Res. 2011;21(8):1273-83. Epub 2011/06/03. doi: 10.1101/gr.122382.111. PubMed PMID: 21632746; PubMed Central PMCID: PMCPmc3149494.

86. Roh TY, Cuddapah S, Zhao K. Active chromatin domains are defined by acetylation islands revealed by genome-wide mapping. Genes Dev. 2005;19(5):542-52. Epub 2005/02/12. doi: 10.1101/gad.1272505. PubMed PMID: 15706033; PubMed Central PMCID: PMCPMC551575.

87. Roh TY, Wei G, Farrell CM, Zhao K. Genome-wide prediction of conserved and nonconserved enhancers by histone acetylation patterns. Genome Res. 2007;17(1):74-81. Epub 2006/12/01. doi: 10.1101/gr.5767907. PubMed PMID: 17135569; PubMed Central PMCID: PMCPMC1716270.

88. Heintzman ND, Stuart RK, Hon G, Fu Y, Ching CW, Hawkins RD, Barrera LO, Van Calcar S, Qu C, Ching KA, Wang W, Weng Z, Green RD, Crawford GE, Ren B. Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. Nat Genet. 2007;39(3):311-8. Epub 2007/02/06. doi: 10.1038/ng1966. PubMed PMID: 17277777.

89. Rada-Iglesias A, Bajpai R, Swigut T, Brugmann SA, Flynn RA, Wysocka J. A unique chromatin signature uncovers early developmental enhancers in humans. Nature. 2011;470(7333):279-83. Epub 2010/12/17. doi: 10.1038/nature09692. PubMed PMID: 21160473; PubMed Central PMCID: PMCPmc4445674.

90. Simpson J, M VANM, Herrera-Estrella L. Photosynthesis-associated gene families: differences in response to tissue-specific and environmental factors. Science. 1986;233(4759):34-8. Epub 1986/07/04. doi: 10.1126/science.233.4759.34. PubMed PMID: 17812887.

91. Clark RM, Wagler TN, Quijada P, Doebley J. A distant upstream enhancer at the maize domestication gene tb1 has pleiotropic effects on plant and inflorescent architecture. Nat Genet. 2006;38(5):594-7. Epub 2006/04/28. doi: 10.1038/ng1784. PubMed PMID: 16642024.

92. Studer A, Zhao Q, Ross-Ibarra J, Doebley J. Identification of a functional transposon insertion in the maize domestication gene tb1. Nat Genet. 2011;43(11):1160-3. Epub 2011/09/29. doi: 10.1038/ng.942. PubMed PMID: 21946354; PubMed Central PMCID: PMCPMC3686474.

93. Timko MP, Kausch AP, Castresana C, Fassler J, Herrera-Estrella L, Van den Broeck G, Van Montagu M, Schell J, Cashmore AR. Light regulation of plant gene expression by an upstream enhancer-like element. Nature. 1985;318(6046):579-82. Epub 1985/12/12. doi: 10.1038/318579a0. PubMed PMID: 3865055.

94. Green PJ, Kay SA, Chua NH. Sequence-specific interactions of a pea nuclear factor with light-responsive elements upstream of the rbcS-3A gene. EMBO J. 1987;6(9):2543-9. Epub 1987/09/01. PubMed PMID: 3678200; PubMed Central PMCID: PMCPMC553672.

95. Valles MB, J; Azorin, F; Puigdomenech, P. Nuclease sensitivity of a maize HRGP gene in chromatin and in naked DNA. Plant Sci. 1991;78(2):225-30.

96. Sullivan AM, Arsovski AA, Lempe J, Bubb KL, Weirauch MT, Sabo PJ, Sandstrom R, Thurman RE, Neph S, Reynolds AP, Stergachis AB, Vernot B, Johnson AK, Haugen E, Sullivan ST, Thompson A, Neri FV, 3rd, Weaver M, Diegel M, Mnaimneh S, Yang A, Hughes TR, Nemhauser JL, Queitsch C, Stamatoyannopoulos JA. Mapping and dynamics of regulatory DNA and transcription factor networks in A. thaliana. Cell reports. 2014;8(6):2015-30. Epub 2014/09/16. doi: 10.1016/j.celrep.2014.08.019. PubMed PMID: 25220462.

97. Zhang W, Zhang T, Wu Y, Jiang J. Genome-wide identification of regulatory DNA elements and protein-binding footprints using signatures of open chromatin in Arabidopsis. Plant Cell.

2012;24(7):2719-31. Epub 2012/07/10. doi: 10.1105/tpc.112.098061. PubMed PMID: 22773751; PubMed Central PMCID: PMCPMC3426110.

98. Zhang W, Wu Y, Schnable JC, Zeng Z, Freeling M, Crawford GE, Jiang J. High-resolution mapping of open chromatin in the rice genome. Genome Res. 2012;22(1):151-62. Epub 2011/11/24. doi: 10.1101/gr.131342.111. PubMed PMID: 22110044; PubMed Central PMCID: PMCPMC3246202.

99. Rodgers-Melnick E, Vera DL, Bass HW, Buckler ES. Open chromatin reveals the functional maize genome. Proc Natl Acad Sci U S A. 2016;113(22):E3177-84. Epub 2016/05/18. doi: 10.1073/pnas.1525244113. PubMed PMID: 27185945; PubMed Central PMCID: PMCPMC4896728.

100. Chatterjee S, Ahituv N. Gene Regulatory Elements, Major Drivers of Human Disease. Annual review of genomics and human genetics. 2017;18:45-63. Epub 2017/04/13. doi: 10.1146/annurev-genom-091416-035537. PubMed PMID: 28399667.

101. Ernst J, Kellis M. Discovery and characterization of chromatin states for systematic annotation of the human genome. Nat Biotechnol. 2010;28(8):817-25. Epub 2010/07/27. doi: 10.1038/nbt.1662. PubMed PMID: 20657582; PubMed Central PMCID: PMCPMC2919626.

102. Deal RB, Henikoff S. A simple method for gene expression and chromatin profiling of individual cell types within a tissue. Dev Cell. 2010;18(6):1030-40. Epub 2010/07/16. doi: 10.1016/j.devcel.2010.05.013. PubMed PMID: 20627084; PubMed Central PMCID: PMCPMC2905389.

103. Furner IJ, Matzke M. Methylation and demethylation of the Arabidopsis genome. Curr Opin Plant Biol. 2011;14(2):137-41. Epub 2010/12/17. doi: 10.1016/j.pbi.2010.11.004. PubMed PMID: 21159546.

104. Meyer P. DNA methylation systems and targets in plants. FEBS Lett. 2011;585(13):2008-15. Epub 2010/08/24. doi: 10.1016/j.febslet.2010.08.017. PubMed PMID: 20727353.

105. Backstrom S, Elfving N, Nilsson R, Wingsle G, Bjorklund S. Purification of a plant mediator from Arabidopsis thaliana identifies PFT1 as the Med25 subunit. Mol Cell. 2007;26(5):717-29. Epub 2007/06/15. doi: 10.1016/j.molcel.2007.05.007. PubMed PMID: 17560376.

106. Yu CP, Chen SC, Chang YM, Liu WY, Lin HH, Lin JJ, Chen HJ, Lu YJ, Wu YH, Lu MY, Lu CH, Shih AC, Ku MS, Shiu SH, Wu SH, Li WH. Transcriptome dynamics of developing maize leaves and genomewide prediction of cis elements and their cognate transcription factors. Proc Natl Acad Sci U S A. 2015;112(19):E2477-86. Epub 2015/04/29. doi: 10.1073/pnas.1500605112. PubMed PMID: 25918418; PubMed Central PMCID: PMCPMC4434728.

107. Bolduc N, Yilmaz A, Mejia-Guerra MK, Morohashi K, O'Connor D, Grotewold E, Hake S. Unraveling the KNOTTED1 regulatory network in maize meristems. Genes Dev. 2012;26(15):1685-90. Epub 2012/08/03. doi: 10.1101/gad.193433.112. PubMed PMID: 22855831; PubMed Central PMCID: PMCPMC3418586.

108. Mathelier A, Zhao X, Zhang AW, Parcy F, Worsley-Hunt R, Arenillas DJ, Buchman S, Chen CY, Chou A, Ienasescu H, Lim J, Shyr C, Tan G, Zhou M, Lenhard B, Sandelin A, Wasserman WW. JASPAR 2014: an extensively expanded and updated open-access database of transcription factor binding profiles. Nucleic Acids Res. 2014;42(Database issue):D142-7. Epub 2013/11/07. doi: 10.1093/nar/gkt997. PubMed PMID: 24194598; PubMed Central PMCID: PMCPMC3965086.

109. O'Malley RC, Huang SC, Song L, Lewsey MG, Bartlett A, Nery JR, Galli M, Gallavotti A, Ecker JR. Cistrome and Epicistrome Features Shape the Regulatory DNA Landscape. Cell. 2016;165(5):1280-92. Epub 2016/05/21. doi: 10.1016/j.cell.2016.04.038. PubMed PMID: 27203113; PubMed Central PMCID: PMCPMC4907330.

110. Yan W, Chen D, Schumacher J, Durantini D, Engelhorn J, Chen M, Carles CC, Kaufmann K. Dynamic control of enhancer activity drives stage-specific gene expression during flower morphogenesis. Nature communications. 2019;10(1):1705. Epub 2019/04/14. doi: 10.1038/s41467-019-09513-2. PubMed PMID: 30979870; PubMed Central PMCID: PMCPMC6461659.

111. Ricci WA, Lu Z, Ji L, Marand AP, Ethridge CL, Murphy NG, Noshay JM, Galli M, Mejia-Guerra MK, Colome-Tatche M, Johannes F, Rowley MJ, Corces VG, Zhai J, Scanlon MJ, Buckler ES, Gallavotti A, Springer NM, Schmitz RJ, Zhang X. Widespread long-range cis-regulatory elements in

the maize genome. Nat Plants. 2019;5(12):1237-49. Epub 2019/11/20. doi: 10.1038/s41477-019-0547-0. PubMed PMID: 31740773; PubMed Central PMCID: PMCPMC6904520.

112. Lu Z, Marand AP, Ricci WA, Ethridge CL, Zhang X, Schmitz RJ. The prevalence, evolution and chromatin signatures of plant regulatory elements. Nat Plants. 2019;5(12):1250-9. Epub 2019/11/20. doi: 10.1038/s41477-019-0548-z. PubMed PMID: 31740772.

CHAPTER 2: IDENTIFICATION OF OPEN CHROMATIN REGIONS IN PLANT GENOMES USING ATAC-SEQ

Marko Bajic, Kelsey A. Maher, and Roger B. Deal

This work is published in *Methods in Molecular Biology* (2018) 1675:183-201. doi: 10.1007/978-1-4939-7318-7_12. Kelsey A. Maher carried out the INTACT-ATAC-seq experiments for the root hair, non-hair, and root tip tissue, the accompanying data processing, and edits for the text of the final manuscript.

ABSTRACT

Identifying and characterizing highly accessible chromatin regions assists in determining the location of genomic regulatory elements and understanding transcriptional regulation. In this chapter we describe an approach to map accessible chromatin features in plants using the Assay for Transposase Accessible Chromatin, combined with high throughput sequencing (ATAC-seq), which was originally developed for cultured animal cells. This technique utilizes a hyperactive Tn5 transposase to cause DNA cleavage and simultaneous insertion of sequencing adapters into open chromatin regions of the input nuclei. The application of ATAC-seq to plant tissue has been challenging due to the difficulty of isolating nuclei sufficiently free of interfering organellar DNA. Here we present two different approaches to purify plant nuclei for ATAC-seq: the INTACT method (Isolation of Nuclei TAgged in Specific Cell Types) to isolate nuclei from individual cell types of the plant, and tissue lysis followed by sucrose sedimentation to isolate sufficiently pure total nuclei. We provide detailed instructions for transposase treatment of nuclei isolated using either approach, as well as subsequent preparation of ATAC-seq libraries. Sequencing-ready ATAC-seq libraries can be prepared from plant tissue in as little as one day. The procedures described here are optimized for *Arabidopsis thaliana* but can also be applied to other plant species.

Key words: ATAC-seq, INTACT system, chromatin, nucleus, transposition, nucleosome, transcription factor, enhancer

INTRODUCTION

Plants are sessile organisms that must precisely regulate their transcription in response to their environment, as well as for proper development, growth, and homeostasis. Transcription is associated with regions of relatively open chromatin, in which cis-regulatory elements such as enhancers and promoters can recruit transcription factors and RNA polymerase II to transcribe DNA (Li et al., 2007). Binding of transcription factors to DNA generally results in the depletion of nucleosomes, rendering these regions hypersensitive to nucleases. Characterizing such regulatory regions throughout the genome has therefore relied on methods that combine enzymatic digestion of nuclear DNA and high-throughput sequencing, such as microccocal nuclease sequencing (MNase-seq, see Chapter 10) and DNase I Hypersensitivity sequencing (DNase-seq) (Song and Crawford, 2010; Ken, 2005). Alternatively, regulatory regions can be inferred by Chromatin Immunoprecipitation sequencing (ChIP-seq, see Chapter 5) where antibodies are used to pull down transcription factors or histone marks associated with active transcription (Park, 2009).

An improved method for identifying accessible regions of chromatin and transcription factor binding is the Assay for Transposase-Accessible Chromatin with high-throughput sequencing (ATAC-seq) (Buenrostro et al., 2013; Buenrostro et a., 2015). This method uses a hyperactive Tn5 transposase to integrate preloaded sequencing adapters into regions of open chromatin (Fig 2.1A). ATAC-seq is a fast protocol with simple library amplification steps and requires very small amounts of starting material, making it a vast improvement over alternative methods. However, a drawback of this protocol is that the hyperactive Tn5 transposase also targets sources of extranuclear genetic material, including the genomes of mitochondria and chloroplasts. This decreases the proportion of reads that map to the nuclear genome, reducing the amount of information that can be used to identify regulatory regions of open chromatin. Such extranuclear reads must be discarded at the start of the data analysis process, diminishing the efficiency of the assay both in terms of cost and in effective use of materials. To gain the maximum efficiency of this powerful procedure, input material free from extranuclear genetic material, such as purified nuclei, is the ideal input for ATAC-seq

In this chapter, we describe the use of two different methods to isolate either total nuclei from tissues or nuclei from specific cell types of *Arabidopsis thaliana* (Fig 2.1B). To isolate total nuclei from plant tissue we use extraction buffers with a non-ionic detergent to lyse organelles, followed by sucrose sedimentation to further purify the nuclei (Gendre et al., 2005). This method of nuclei isolation can be done in any lab on most plant tissues. However, these partially purified nuclei still contain some organellar DNA in addition to nuclear DNA, which reduces the efficiency of Tn5 transposition to nuclear DNA and results in fewer sequencing reads that map to nuclear DNA. In addition, we describe the Isolation of Nuclei TAgged in specific Cell Types (INTACT) method to isolate nuclei from tissue or from specific cell types (Deal and Henikoff, 2010). This system uses two transgenes for nuclear targeting for affinity purification: 1) the Nuclear Tagging Fusion (NTF) construct, which encodes a fusion of WPP nuclear envelope-targeting domain, a Green Fluorescent Protein (GFP), and the Biotin Ligase Recognition Peptide (BLRP); and 2) an *E. coli* biotin ligase (BirA), which biotinylates the BLRP tag. The BirA is expressed from a constitutive promoter while the NTF is expressed either from a constitutive or cell type-specific promoter. The specificity of the NTF promoter determines which cell types will have biotinylated nuclei and can then be isolated by affinity purification with streptavidin-coated magnetic beads (Wang and Deal, 2015). A key advantage of the INTACT approach is not only that the isolated nuclei have less organellar DNA contamination, but also that this method can be used to selectively isolate nuclei from specific cell types. While INTACT is a powerful technique, it does require that stable transgenic lines containing BirA and NTF cassettes for the cell type of interest are available, which are time-consuming to generate and can be limiting for many species. Even so, the protocol described here, particularly ATAC-seq using sucrose sedimentation-purified nuclei, can readily be adapted for chromatin profiling in any plant species.

MATERIALS

2.1 Equipment

1. Porcelain 50 mL mortar and pestle, or equivalent.

2. Liquid nitrogen.
3. Metal lab spoon.
4. DynaMag 2 magnetic rack for 1.5 mL tubes (e.g. Life Technologies, catalog no. 12321D).
5. DynaMag 15 magnetic rack for 15 mL tubes (e.g. Life Technologies, catalog no. 12301D).
6. MagWell 96 well magnetic separator plate (e.g. EdgeBio, catalog no. 57624)
7. Nylon cell strainers with 70 μm pores.
8. Long-stem analytical funnel.
9. Pipet-Aid.
10. Sterile 10 mL plastic serological pipettes.
11. Eppendorf tubes, 1.5 mL.
12. PCR tubes, 0.2 mL.
13. Falcon tubes, 15 and 50 mL.
14. Nutator platform rotator.
15. Hemocytometer (e.g. Hausser Bright Line hemocytometer, Fisher Scientific)
16. Microcentrifuge and refrigerated centrifuge with rotor for 15 mL tubes.
17. Cold room, 4 °C.
18. Molecular biology grade water.
19. Sterile disposable filter unit, 500 mL.
20. Sterile 0.2 μm syringe filter.
21. Sterile 10 mL plastic syringe.
22. Thermal cycler
23. Real-Time PCR machine
24. A 64-bit computer with at least 1 TB hard disk and 16 Gb of memory for ATAC-seq data analysis.
25. Fluorescent microscope.

2.2 Stock Solutions and Reagents

1. Complete, EDTA-free Protease Inhibitors (e.g. Roche).

2. Stock solution of 2 M spermidine. Prepare by dissolving 2.904 g spermidine powder in 10 mL water. Aliquot 1mL of solution per 1.5 mL Eppendorf tube and store at -20 °C.

3. Stock solution of 200 mM spermine. Prepare by dissolving 0.4047 g spermine powder in 10 mL of water. Aliquot 1 mL of solution per 1.5 mL Eppendorf tube and store at -20 °C.

4. Stock solution of incomplete Nuclei Purification Buffer (NPBi): 20 mM MOPS, 40 mM NaCl, 90 mM KCl, 2 mM EDTA, 0.5 mM EGTA, adjusted to pH 7 with 2M KOH. Filter sterilize the solution and degas under vacuum for 10 minutes. Store at 4 °C for up to 3 months.

5. Stock solution of 10% Triton X-100.

6. Stock solution of 10X DAPI. Prepare by dissolving 10 mg DAPI powder in 5 mL water, for a final concentration of 2 μg/μL. Filter sterilize the solution and store at 4 °C in the dark for several months. To stain nuclei with DAPI, dilute the 10X DAPI solution to 1X using water (final concentration of 0.2 μg/μL), and use within 2-3 hours.

2.3 Purification of Tagged Nuclei using INTACT

1. Plant material: tissue from transgenic plants expressing both NTF and BirA in the cell type of interest. INTACT transgenic lines targeting the root epidermal hair and non-hair cell types, as well as INTACT plasmid vectors are available from the Arabidopsis Biological Resource Center at Ohio State University.

2. M-280 Streptavidin Dynabeads (e.g. Life Technologies).

3. Nuclei Purification Buffer (NPB): 20 mM, 40 mM NaCl, 90 mM KCl, 2 mM EDTA, 0.5 mM EGTA, 0.5 mM spermidine, 0.2 mM spermine, 1X Roche Complete protease inhibitors, adjusted to pH 7 with 2M KOH. Prepare by adding spermidine, spermine, and Roche Complete protease inhibitors to NPBi just before starting the INTACT nuclei purification procedure. Keep solution on ice, and use within 1 hour of preparation.

4. Nuclei Purification Buffer containing 0.1% Triton X-100 (NPBt): 20 mM MOPS pH 7, 40 mM NaCl, 90 mM KCl, 2 mM EDTA, 0.5 mM EGTA, 0.5 mM spermidine, 0.2 mM spermine, 0.1% (v/v) Triton X-100. Prepare by adding spermidine, spermine, and Triton X-100 to NPBi just before starting the INTACT nuclei purification procedure. Keep solution on ice, and use within 1 day of preparation.

2.4 Purification of Total Nuclei using Sucrose Sedimentation

1. Plant material: fresh or frozen plant tissue.
2. Stock solution of 1M Tris-HCl pH 8
3. Stock solution of 1M MgCl2
4. Stock solution of 2M sucrose.
5. Nuclei Purification Buffer (NPB): 20 mM MOPS pH7, 40 mM NaCl, 90 mM KCl, 2 mM EDTA, 0.5 mM EGTA, 0.5 mM spermidine, 0.2 mM spermine, 1X Roche Complete protease inhibitors. Prepare by adding spermidine, spermine, and Roche Complete protease inhibitors to NPBi just before starting the nuclei purification procedure. Keep solution on ice, and use within 1 hour of preparation.
6. Nuclei Extraction Buffer 2 (NEB 2): 0.25 M Sucrose, 10 mM Tris-HCl pH 8, 10 mM MgCl2, 1% Triton X-100, 1X Roche Complete Protease Inhibitors. Prepare solution just before use, keep on ice, and use within 1 hour of preparation.
7. Nuclei Extraction Buffer 3 (NEB 3): 1.7 M Sucrose, 10 mM Tris-HCl pH 8, 2 mM MgCl2, 0.15% Triton X-100, 1X Roche Complete Protease Inhibitors. Prepare solution just before use, keep on ice, and use within 1 hour of preparation.

2.5 Tagmentation of Chromatin by Tn5 transposase

1. Nextera Library Kit (Illumina, FC-121-1030).
2. MinElute PCR Purification kit (Qiagen).

2.6 Sequencing Library Preparation

1. ATAC Primer 1

(AATGATACGGCGACCACCGAGATCTACACTCGTCGGCAGCGTCAGATGTG)

2. ATAC barcoded Primer 2

(CAAGCAGAAGACGGCATACGAGATNNNNNNNNGTCTCGTGGGCTCGGAGATGT);

N's indicate the 8-base index sequence. Each library to be pooled for sequencing should be amplified with a different barcoded primer 2. See Supplementary Table 2.1 of (Buenrostro et al., 2013) for all primer sequences.

3. NEBNext High-Fidelity 2X PCR Master Mix (NEB).
4. Solution of 20X EvaGreen dye (Biotium).
5. Solution of 50X ROX dye (Invitrogen).
6. MinElute PCR Purification kit (Qiagen).
7. Agencourt Ampure XP PCR Purification beads (Beckman Coulter).
8. 100% ethanol.
9. Horizontal electrophoresis gel box and power source.
10. 302 nm ultraviolet transilluminator.
11. NEBNext Library Quantification kit for Illumina (NEB)

METHODS

Users should either begin at section 3.1 for affinity purification of nuclei using INTACT, or at section 3.2 for isolation of total nuclei. In either case, the purified nuclei are used for tagmentation by Tn5 transposase in step 3.3. All procedures are carried out at room temperature (25 °C) unless otherwise specified.

3.1 Purification of Tagged Nuclei Using INTACT

1. Grind tissue (3 g of roots or 0.5 g of leaves) to a fine powder in liquid nitrogen using a mortar and pestle. Using a nitrogen-cooled metal lab spoon, quickly transfer the frozen tissue powder to another mortar

containing 10 mL of ice-cold Nuclei Purification Buffer (NPB). Thoroughly resuspend the powder in NPB by grinding it with a new, cold pestle (see Note 1).

2. Use a 10 mL serological pipette to draw up the tissue suspension and filter it through a 70 μm nylon cell strainer, placed in the center of a long-stemmed funnel. Collect the flow-through in a chilled 15 mL tube on ice.

3. Spin down the nuclei at 1,200 x g for 10 minutes at 4 °C. Use a 10 mL serological pipet and then a 1 mL pipette tip to carefully remove as much of the supernatant as possible without disturbing the pellet.

4. Gently resuspend the pellet in 1 mL of ice-cold NPB. Transfer the crude nuclei suspension to a 1.5 mL tube. Keep on ice.

5. Wash the appropriate amount of Streptavidin M280 Dynabead suspension (25 μL for nuclei from 3 g of roots or 10 μL for 0.5 g of leaves) with 1 mL of ice-cold NPB in a 1.5 mL tube. Collect the beads on the DynaMag2 magnetic rack. Discard the supernatant and resuspend the beads with ice-cold NPB to their original volume (e.g. 25 μL). Keep on ice.

6. Add the washed and resuspended beads to the 1 mL of resuspended nuclei from Step 4. Rotate on a nutator in a 4 °C cold room for 30 minutes. Work in the 4 °C cold room for Steps 7-14.

7. Transfer the 1 mL bead-nuclei mixture to a 15 mL tube and slowly add to it 13 mL of ice-cold NPBt. Mix gently and place on a nutator for 30 seconds.

8. Place the 15 mL tube in the DynaMag 15 magnetic rack for 2 minutes to capture the nuclei-beads along the walls of the tube.

9. Slowly remove the NPBt supernatant with a serological pipette, making sure not to disturb the beads on the side walls of the tube. Gently resuspend the beads with 14 mL of ice-cold NPBt, mix gently, and place on a nutator for 30 seconds.

10. Place the 15 mL tube in the DynaMag 15 magnetic rack for 2 minutes to capture the nuclei and beads.

11. Repeat Steps 9 and 10 one more time, for a total of three washes.

12. Slowly remove the NPBt supernatant with a serological pipette. Resuspend the beads in 1 mL of ice-cold NPBt. Remove 25 μL of this nuclei-bead suspension to a 0.6 ml tube on ice for counting captured nuclei with a hemocytometer.

13. Transfer the remaining nuclei-bead suspension to an ice-cold 1.5 mL tube. Place the 1.5 mL tube in the DynaMag 2 magnetic rack to capture the beads along the walls of the tube.

14. Carefully remove the NPBt supernatant and resuspend the nuclei-beads in 20 μL of ice-cold NPB. Keep on ice until the nuclei are counted and ready for tagmentation. (see Note 2).

15. To view and quantify nuclei under a light microscope, add 1 μL of diluted DAPI solution (0.2 μg/μL) to each 25 μL aliquot of nuclei from Step 12. Mix well, and place on ice for 5 minutes in the dark.

16. Use a hemocytometer to count the DAPI-stained, bead-bound nuclei and determine the total yield. Purified nuclei should appear as shown in Figure 2.1C (see Note 3).

17. Use the calculated total yield to determine the volume of resuspended nuclei from Step 14 needed to obtain 50,000 nuclei for the ATAC-seq reaction. Transfer this volume of resuspended nuclei to a new 0.2 mL tube, and keep on ice. Immediately proceed to Section 3.3.

3.2 Purification of Total Nuclei Using Sucrose Sedimentation

1. Grind 0.1 to 1 g of plant tissue to a fine powder in liquid nitrogen using a mortar and pestle (see Note 4).

2. Using a nitrogen-cooled metal lab spoon, quickly transfer the frozen tissue powder to another mortar containing 10 mL ice-cold NPB. Thoroughly resuspend the powder in NPB by grinding it with a new, cold pestle.

3. Use a 10 mL serological pipette to draw up the tissue suspension and filter it through a 70 μm nylon cell strainer, placed in the center of a long-stemmed funnel. Collect the flow-through into a 15 mL tube on ice.

4. Centrifuge the tube at 1,200 x g for 10 minutes at 4 °C.

5. Gently remove the supernatant and gently but thoroughly resuspend the pellet in 1 mL of ice-cold NEB2 buffer. Transfer this suspension to a new 1.5 mL microcentrifuge tube.

6. Spin the resuspended nuclei at 12,000 x g for 10 minutes at 4 °C.

7. Carefully remove the supernatant and resuspend the pellet thoroughly in 300 µL of NEB3 buffer.

8. Add 300 µL of ice-cold NEB3 to a new 1.5 mL microcentrifuge tube. Carefully layer the resuspended pellet from Step 7 on top of the fresh NEB3. Centrifuge at 16,000 x g for 10 minutes at 4 °C (see Note 5).

9. Carefully remove the supernatant and resuspend the nuclei pellet in 1 mL of cold NPB. Keep these nuclei on ice.

10. Remove 25 µL of this nuclei suspension and move to a fresh 0.6 ml tube on ice. To this add 1 µL of diluted DAPI solution (0.2 µg/µL). Mix well and place on ice for 5 minutes in the dark.

11. Use a hemocytometer to quantify the DAPI-stained nuclei and determine the total yield. Purified nuclei should appear as shown in Fig 2.1C (see Note 6).

12. Use the calculated total yield to determine the volume of resuspended nuclei from Step 9 needed to obtain 50,000 nuclei for the ATAC-seq reaction. Transfer this volume of the resuspended nuclei to a new 0.2 mL tube, and keep on ice. Immediately proceed to Section 3.3.

3.3 Tagmentation with Tn5 Transposase

1. Prepare the transposition reaction master mix in a 0.2 mL PCR tube on ice according to Table 2.1 and mix well. The volumes given in Table 2.1 are for a single reaction with 50,000 nuclei.

2. If the nuclei were isolated using the Sucrose Sedimentation procedure, pellet 50,000 nuclei from Subheading 3.2 Step 9 by spinning the appropriate volume of nuclei at 1,500 x g for 7 minutes at 4 °C. Remove the supernatant, and resuspend the nuclei in 50 µL of ice-cold transposition reaction mix prepared in step 1. Move the reaction to a 0.2 mL PCR tube on ice. If the nuclei were isolated using the INTACT procedure, move 50,000 bead-bound nuclei from Subheading 3.1 Step 14 into a 0.2 mL tube and capture

the beads on the tube wall in a MagWell 96 well magnetic plate on ice. Remove the supernatant, and resuspend the bead-bound nuclei in 50 µL of ice-cold transposition reaction mix. Keep on ice.

3. Place the transposition reaction in a thermal cycler block pre-warmed to 37 °C and incubate for 30 minutes with occasional gentle mixing to keep the nuclei in suspension.

4. Purify the transposed DNA using the Qiagen MinElute PCR purification kit according the manufacturer's instructions. Elute DNA in 11 µL of elution buffer EB, provided in the kit. DNA can now be stored at -20 °C until future use, or used immediately for PCR amplification.

3.4 PCR Amplification of the DNA Library

1. Prepare the PCR amplification mix in a 0.2 mL tube on ice according to Table 2.2. Mix well, and perform PCR cycling as described in Table 2.3 (see Note 7).

2. Once the thermal cycler reaches 4 °C, remove the samples and place them on ice.

3. To determine the number of additional PCR cycles needed to adequately amplify the DNA library, prepare the qPCR Library Amplification Mix described in Table 2.4 in a 0.2 mL PCR tube. Keep the mixture on ice.

4. Perform thermal cycling in the qPCR machine according to Table 2.5.

5. To determine the optimal number of cycles needed to amplify the remaining 45 µL of each library from Step 2, view the linear fluorescence versus cycle number plot on the qPCR machine once the reaction is finished. The cycle number at which the fluorescence for a given reaction is at 1/3 of its maximum is the number of additional cycles (N) that each library requires for adequate amplification (see Note 8).

6. Run the remaining 45 µL of each PCR reaction from Step 2 according to Table 2.6.

7. Purify the libraries by mixing Ampure XP beads with the reaction products at a 1.5:1 ratio of beads:PCR sample by volume (see Note 9). Incubate at room temperature for 5 minutes.

8. Place the 0.2 mL tube on the MagWell 96 well magnetic plate for 1 minute to capture the Ampure beads, and discard the supernatant.

9. With the tube still in the magnetic plate, wash the beads twice for 30 seconds each with 200 µL of 80% ethanol without disturbing the bead pellet. After the last wash, allow the beads to dry for 5 minutes to remove all traces of ethanol (see Note 10).

10. Remove the tube from the magnet and resuspend the bead pellet in 20 µL 10 mM Tris pH 8. Incubate at room temperature for 2 minutes, capture the beads on the magnet, and transfer the supernatant into a fresh 0.2 mL PCR tube on ice. A small aliquot of the library, 1-2 µL, can be run on a 2% agarose gel to visualize the abundance and size distribution of amplified libraries (Fig 2.2A) (see Note 11). The purified libraries can now be stored at -20 °C.

11. Quantify the molar concentrations of the libraries using the NEBNext Library Quantification kit for Illumina, according to manufacturer's directions. Alternatively, other qPCR-based library quantification kits can be used to determine the concentration of the amplified libraries.

12. Once quantified, the libraries are ready for pooling and high-throughput sequencing on the Illumina platform (see Note 12).

13. The quality of the sequencing reads, alignment to the genome, fragment size distribution (Fig 2.2B), and downstream analyses can be performed as described in Note 13. A genome browser shot of the typical Arabidopsis ATAC-seq data from libraries made using the procedures described here can be seen in Fig 2.2C.

NOTES

1. This protocol is optimized for 3 g of root or 0.5 g of leaf tissue from *Arabidopsis thaliana*. Ground leaf tissue contains more debris, relative to roots, and therefore requires a lower amount of starting material to obtain highly purified nuclei. INTACT may also be performed on fresh tissue by chopping the tissue in NPB as opposed to grinding to a fine powder using liquid nitrogen. However, this approach does require the use of fresh tissue. The number of samples that can be run through INTACT purification simultaneously is mainly limited by the capacity of the DynaMag 15 magnetic rack used for nuclei capture. Up to four separate samples can be processed in parallel using one DynaMag 15 magnetic rack.

Using an INTACT line with nuclei labeled in the root epidermal non-hair cell type, approximately 200,000 purified nuclei can be obtained from 3g of roots. Larger amounts of tissue can be used for purifying nuclei from less abundant cell types, and this generally only requires adjustments to the amount of streptavidin beads used and the volume of solution used for bead capture. See (Wang and Deal, 2015) for more details on variations in the INTACT procedure.

2. After isolating the bead bound nuclei, keep the sample on ice while quantifying the nuclei from the aliquot in Subheading 3.1 Step 12. Do not freeze the isolated nuclei before doing tagmentation and library preparation. Freezing and thawing of isolated nuclei can disrupt protein-DNA interactions.

3. After DAPI staining, nuclei purified by INTACT can be easily identified and counted using a hemocytometer. The ideal setup for visualizing nuclei is under a mix of dim white light and DAPI channel fluorescence. The dim white light allows for visualization of the hemocytometer grid and the beads, and the DAPI fluorescence allows for the visualization of nuclei. A sample image of isolated bead-bound nuclei is shown in Fig 2.1C. A nucleus is identified as a circle that fluoresces in the DAPI channel and has several beads clustered around it. Minimal cellular debris or contaminating unbound nuclei should be observed in the final product. These contaminants may be further reduced by using fewer beads and by increasing the volumes of NPB and NPBt used during purification as described in Note 1.
We have successfully used as few as 20,000 to as many as 200,000 INTACT-purified nuclei in this procedure without altering any other parameters of the protocol presented here.

4. This protocol is optimized for less than 1 g of root or 0.5 g of leaf tissue. Ground leaf tissue contains more debris relative to roots, and therefore requires a lower amount of starting material to obtain purified nuclei. As with the INTACT protocol, sucrose sedimentation of nuclei may also be performed on fresh tissue by chopping the tissue in NPB as opposed to grinding to a fine powder using liquid nitrogen.

However, this approach does require the use of fresh tissue. We recommend starting with the minimum amount of tissue needed to obtain the required number of nuclei (e.g. 50,000 per ATAC-seq reaction).

5. Proper separation of nuclei from other cellular debris requires the nuclei to pass through the sucrose cushion during centrifugation. The NEB3 resuspended nuclei should therefore be placed gently on top of NEB3 layer present in the tube. After centrifugation, the contaminating organelles and debris may be visible at the top of the tube. If leaf tissue was used, the top layer will become greener after centrifugation and the pellet will become noticeably less green than it was prior to centrifugation.

6. After DAPI staining, nuclei purified by sucrose sedimentation can be identified and quantified using a hemocytometer. A mixture of DAPI-channel fluorescence and white light illumination allows the stained nuclei and the hemocytometer grid to be seen simultaneously. A sample image of isolated nuclei is shown in Fig 2.1C. A nucleus is identified as a punctate circle with strong DAPI fluorescence. The nucleus is typically ~5 μm in size and can be easily identified at 200X and 400X magnifications. Cellular debris may be observed in the final preparation, but this generally does not affect the outcome of the ATAC-seq procedure. To reduce cellular debris contamination, starting tissue can be chopped with a razor blade (see Note 4) and/or additional NEB3 wash steps may also be done by repeating Subheading 3.2 steps 7-9 for a second sucrose cushion centrifugation.

7. Ensure that all work surfaces, pipettes, and reagents needed for amplification and library preparation are free of DNA contamination. For library amplification, unique barcoded adapters are used for each sample if multiple libraries are to be sequenced in an individual flow cell lane. The sequences of all primers can be found in the supplementary material of (Buenrostro et al., 2013).

8. The number of PCR cycles needed to amplify ATAC libraries is determined by the PCR reaction in Subheading 3.4 step 5. We recommend using the minimum number of cycles necessary to obtain a

sufficient molar amount of library for Illumina sequencing. This must be determined empirically and will also depend on the number of libraries to be pooled for sequencing.

9. The ratio of Ampure XP PCR Purification beads to PCR volume determines the size of purified DNA fragments isolated. The 1.5 Ampure bead to PCR reaction ratio results in the isolation of DNA fragments shown in Fig 2.2A. Using ratios that have higher proportions of beads may result in purification of sequencing adapters and PCR primers, which can negatively affect sequencing.

10. A drying time of 5 minutes is generally sufficient to remove all traces of ethanol from the beads, but this time may vary based on humidity and room temperature. Georgia is very humid in the summer. Ensure that all ethanol has evaporated before moving on to the next step. Do not allow beads to dry to the extent that the pellet begins to crack.

11. Libraries can generally be visualized by agarose gel electrophoresis followed by ethidium bromide staining. Sensitivity can be greatly increased by staining the gel with Sybr green stain or using an Agilent Bioanalyzer or equivalent instrument, if available.

The libraries that we have prepared using this method generally present as a DNA smear starting at ~180 bp and ranging to greater than 1 kb, with peak intensity between ~180 – 500 bp (See Figure 2.2A). The original publication on ATAC-seq (Buenrostro et al., 2013) reported a nucleosome-like periodicity in the library size distribution, but we have not observed this phenomenon as assayed by either electrophoresis or estimation of fragment size distribution based on distance between paired-end sequencing reads, as shown in Fig 2.2B. This lack of observed nucleosome fractions may be due to size selection of library fragments by Ampure XP beads and the low transposase to nuclei ratio described in this protocol.

12. Paired-end sequencing is recommended in order to maximize the number of transposase integration events that can be observed in a given sample and to allow measurement of the length of the sequenced fragments (Fig 2.2B).

To identify open chromatin regions in Arabidopsis, users should aim to obtain at least 10-20 million reads per library that map to the nuclear genome. For transcription factor footprinting the number of nuclear genome-mapping reads should be increased to at least 100 million per library.

When using sucrose sedimentation for nuclei purification, users should expect ~50% of reads to map to the nuclear genome, while the use of INTACT purification will increase this number to > 90%.

13. Sequencing reads are checked for overall quality using FastQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/) or equivalent. The reads are aligned to the TAIR10 Arabidopsis thaliana genome (https://www.arabidopsis.org/download/index-auto.jsp?dir=%2Fdownload_files%2FGenes%2FTAIR10_genome_release) using Bowtie2 (http://bowtie-bio.sourceforge.net/bowtie2/index.shtml). The resulting SAM file is converted to a binary BAM file, which is sorted and indexed using Samtools (http://samtools.sourceforge.net/). The quality of the resulting BAM file, including fragment size distribution, is analyzed using Picard Tools (https://broadinstitute.github.io/picard/). Alignment data is visualized using the Integrated Genome Viewer (http://software.broadinstitute.org/software/igv/). For ease of visualization, BAM files were converted to BigWig files using DeepTools BamPECoverage tool (http://deeptools.readthedocs.io/en/latest/index.html). Downstream analyses of ATAC-seq data include calling peaks with HOMER (http://homer.salk.edu/homer/index.html), editing BED files with bedtools (http://bedtools.readthedocs.io/en/latest/) and identifying transcription factor footprints using pyDNase (http://pythonhosted.org/pyDNase/).

ACKNOWLEDGEMENTS

This work was supported by the National Science Foundation Grant no. 1238243. We thank Paja Sijacic and Shannon Torres for helping to optimize the protocol for nuclei isolation and for suggestions on the manuscript.

TABLES AND FIGURES

Figure 2.1

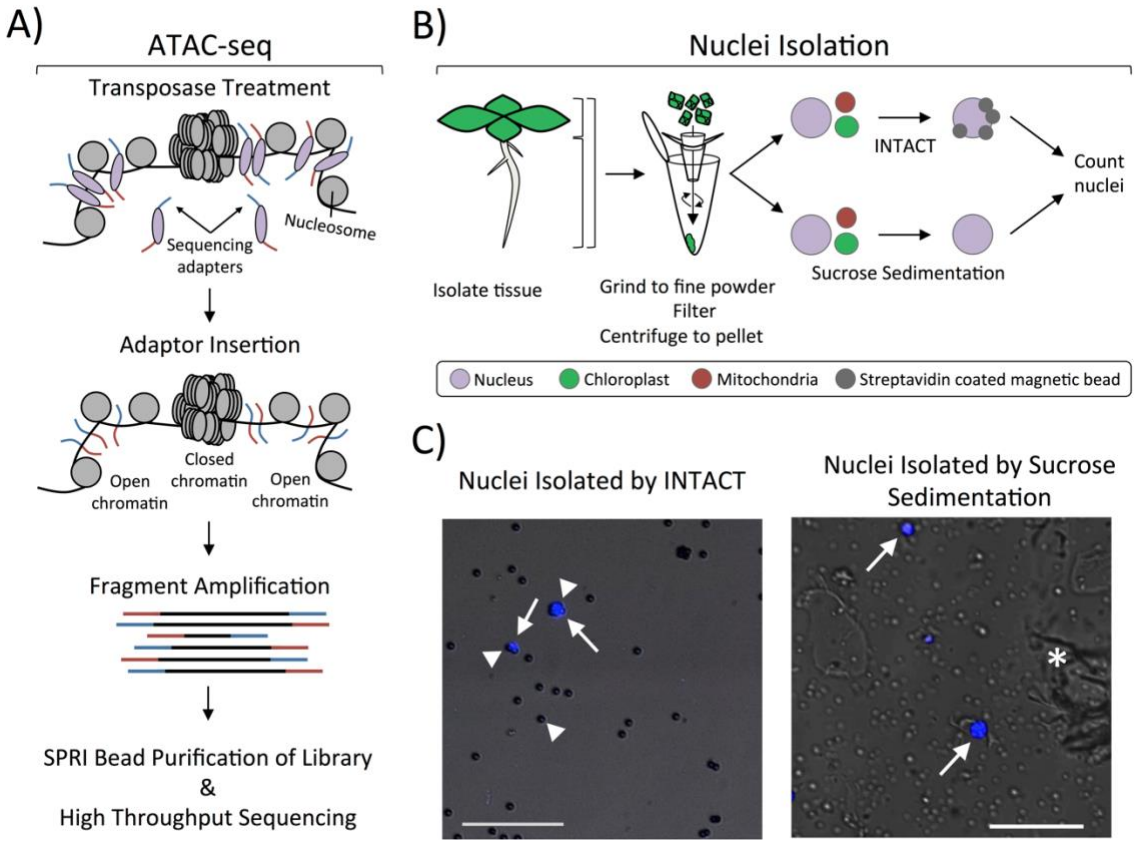


Figure 2.1 ATAC-seq profiling using nuclei isolated by INTACT or sucrose sedimentation. A) Overview of the ATC-seq procedure. Nuclei are incubated with sequencing adapter-loaded Tn5 transposase, which diffuses into the nucleus to interact with chromatin. Sequencing adapters are inserted into open chromatin regions, and the fragmented DNA is amplified wherever the sequencing adapters were inserted. This generates a library of DNA fragments in which each end represents an insertion site. The amplified libraries are purified and sequenced with next generation sequencing. B) Two different methods for purifying nuclei from Arabidopsis thaliana can be used: 1) INTACT for isolating nuclei from specific cell types, and 2) sucrose sedimentation to isolate total nuclei from input tissue. The two methods have the same initial steps: tissue is collected from a specific part of the plant (root, leaf, or the entire plant), ground to a fine powder, resuspended, filtered, and centrifuged to pellet nuclei and cellular debris. Nuclei isolation using tissue that expresses INTACT transgenes uses streptavidin coated magnetic beads to affinity purify

biotinylated nuclei out of the resuspended pellet. This allows for the isolation of nuclei from specific cell-types that express the nuclear tagging fusion (NTF) and the biotin ligase BirA, resulting in very low contamination by organellar genomes. Alternatively, total nuclei can be isolated from tissue by resuspending the nuclei/debris pellet in a buffer with Triton X-100 to lyse organelles and centrifuging through a dense sucrose layer. Nuclei isolated from both procedures are stained with DAPI and quantified using a hemocytometer. C) Fluorescent microscope images of nuclei (white arrows) stained with the DNA-binding dye DAPI (blue) isolated either through INTACT or sucrose sedimentation. INTACT isolated nuclei are identified by their DAPI-fluorescence and binding to multiple beads (white arrowhead). Beads are easily visualized by increasing transmission of white light while viewing the nuclei in the DAPI channel. Sucrose sedimentation isolated nuclei (white arrows) are DAPI-stained objects around 4-6 μm in diameter, although they can vary in size and shape depending on starting tissue. Much more cellular debris (white asterisk) is observed in sucrose sedimentation-isolated nuclei as compared to INTACT-purified nuclei, but this should not impact the procedure described here. Each picture contains a 50 μm scale bar shown at the bottom left.

Figure 2.2

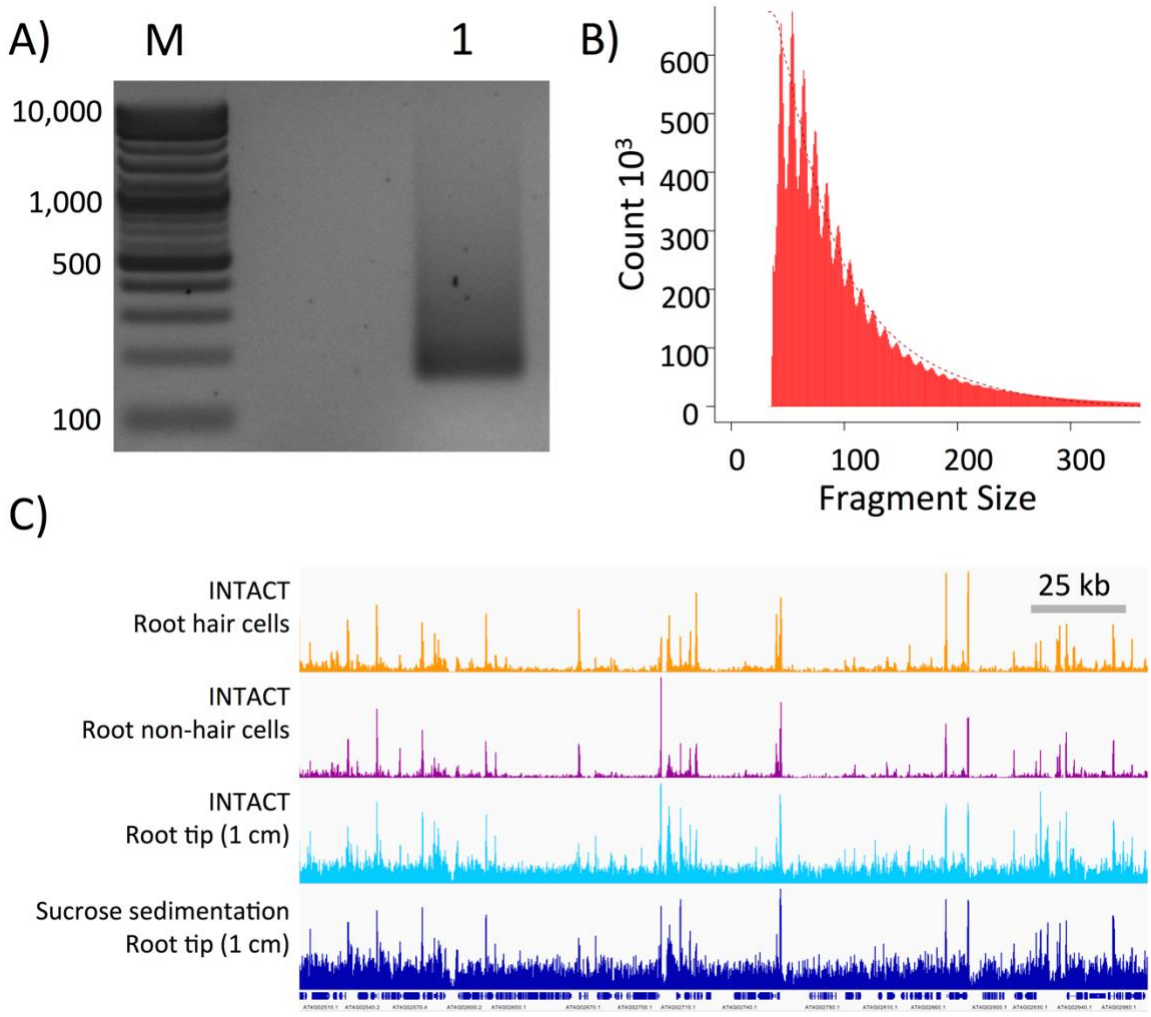


Figure 2.2 ATAC-seq library preparation and high-throughput sequencing. A) An amplified ATAC-seq library purified with Ampure XP beads (lane "1") was resolved in a 2% agarose gel stained with ethidium bromide. Lane "M" is the molecular weight marker lane. Amplified library fragments generally range in size from 180 bp to several kb in size. The size distribution of the resolved gel may vary somewhat, but the final product should be free of adapter dimers (distinct band around 125 bp) and primer dimers (distinct band around 80 bp). See Note 11. B) Insert sizes of ATAC-seq paired-end reads from 50,000 nuclei isolated by INTACT from non-hair cells calculated using the InsertSizeMetrics option from Picard Tools (Note 13). The distribution shows periodicity of helical pitch of DNA for fragments smaller than 200 bp. Fragments containing one or more nucleosomes, related to insert periodicity increasing in 150 bp, were not

observed using the transposase:nuclei and bead:DNA ratios described in this protocol. C) Integrated Genome Viewer snapshot of four different libraries sequenced on the Illumina platform. The tracks shown are of ATAC sequencing reads from INTACT isolated nuclei from root hair cells (orange), root non-hair cells (purple), root tip (cyan), and sucrose sedimentation isolated nuclei from 1 cm root tip (navy). Gene tracks are shown below the ATAC-seq tracks and a 25 kb scale bar is shown.

Table 2.1

Transposition reaction mix

Component	Volume (µL)
2X TD Buffer	25
Water	22.5
TDE1 Transposase	2.5
Total	50

Table 2.2

Transposed DNA Amplification mix

Component	**Volume (µL)**
Transposed DNA (from Subheading 3.3 step 4)	10
Water	10
25 µM ATAC Primer 1	2.5
25 µM ATAC barcoded Primer 2*	2.5
2X NEBNext High Fidelity PCR Mix	25
Total	50

*A different barcoded Primer 2 should be used for each library that is to be pooled into a single sequencing run.

Table 2.3

Thermal Cycling Conditions for Transposed DNA Amplification

Cycle number	Temperature (°C)	Time
1	72	5 min
	98	30 sec
5 cycles	98	10 sec
	63	30 sec
	72	1 min
	4	Hold

Table 2.4

qPCR Library Amplification Mix

Component	Volume (µL)
Amplified library (from Subheading 3.4 step 2)	5
Water	0.45
25 µM ATAC Primer 1	0.5
25 µM ATAC barcoded Primer 2	0.5
20X Evagreen dye	0.75
50X ROX dye*	0.30
2X NEBNext High Fidelity PCR Mix	7.5
Total	15

*ROX concentration may vary depending on qPCR instrument. The amount described here is optimized for the ABI Step-One-Plus instrument.

Table 2.5

qPCR Cycling Conditions to Determine Additional Library Amplification Cycles

Cycle number	Temperature (°C)	Time
1	98	30 sec
20 cycles	98	10 sec
	63	30 sec
	72	1 min

Table 2.6

Final Library Amplification

Cycle number	**Temperature (°C)**	**Time**
1	98	30 sec
N cycles	98	10 sec
	63	30 sec
	72	1 min
	4	Hold

LITERATURE CITED

Li B., Carey M., Workman J.L. (2007) The role of chromatin during transcription. Cell 128:707-719

Song L., Crawford G.E. (2010) DNase-seq: a high-resolution technique for mapping active gene regulatory elements across the genome from mammalian cells. Cold Spring Harb Protoc 2010: pdb prot5384. doi:10.1101/pdb.prot5384

Ken Z. (2005) Micrococcal Nuclease Analysis of Chromatin Structure. Current Protocols in Molecular Biology 21.1:1-17

Park P.J. (2009) ChIP-seq: advantages and challenges of a maturing technology. Nat Rev Genet 10: 669-680

Buenrostro J.D., Giresi P.G., Zaba L.C., Chang H.Y., Greenleaf W.J. (2013) Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. Nat Methods 10: 1213-1218

Buenrostro J.D., Wu B., Chang H.Y., Greenleaf W.J. (2015) ATAC-seq: A method for assaying chromatin accessibility genome-wide. Curr Protoc Mol Biol 109:21.29 1-9

Gendrel A., Lippman Z., Martienssen R., Colot V. (2005) Profiling histone modification patterns in plants using genomic tiling microarrays. Nature Methods 2: 213-218

Deal R.B., Henikoff S. (2010) A simple method for gene expression and chromatin profiling of individual cell types within a tissue. Dev Cell 18:1030-1040

Wang D., Deal R.B. (2015) Epigenome Profiling of Specific Plant Cell Types Using a StreamLined INTACT Protocol and ChIP-seq. Mehods Mol Biol 1284: 3-25

CHAPTER 3: PROFILING OF ACCESSIBLE CHROMATIN REGIONS ACROSS MULTIPLE PLANT SPECIES AND CELL TYPES REVEALS COMMON GENE REGULATORY PRINCIPLES AND NEW CONTROL MODULES

Kelsey A. Maher*, Marko Bajic*, Kaisa Kajala, Mauricio Reynoso, Germain Pauluzzi, Donnelly A. West, Kristina Zumstein, Margaret Woodhouse, Kerry Bubb, Michael W. Dorrity, Christine Queitsch, Julia Bailey-Serres, Neelima Sinha, Siobhan M. Brady, and Roger B. Deal

This work is published in *The Plant Cell* (2018) 1:15-36. doi: 10.1105/tpc.17.00581.

Supplemental tables can be found in the online publication.

*These authors contributed equally to this work. Contribution details can be found in the Author Contributions section.

ABSTRACT

The transcriptional regulatory structure of plant genomes remains poorly defined relative to animals. It is unclear how many *cis*-regulatory elements exist, where these elements lie relative to promoters, and how these features are conserved across plant species. We employed the Assay for Transposase-Accessible Chromatin (ATAC-seq) in four plant species (*Arabidopsis thaliana*, *Medicago truncatula*, *Solanum lycopersicum*, and *Oryza sativa*) to delineate open chromatin regions and transcription factor (TF) binding sites across each genome. Despite 10-fold variation in intergenic space among species, the majority of open chromatin regions lie within 3 kb upstream of a transcription start site in all species. We find a common set of four TFs that appear to regulate conserved gene sets in the root tips of all four species, suggesting that TF-gene networks are generally conserved. Comparative ATAC-seq profiling of *Arabidopsis* root hair and non-hair cell types revealed extensive similarity as well as many cell type-specific differences. Analyzing TF binding sites in differentially accessible regions identified a MYB-driven regulatory module unique to the hair cell, which appears to control both cell fate regulators and abiotic

stress responses. Our analyses revealed common regulatory principles among species and shed light on the mechanisms producing cell type-specific transcriptomes during development.

INTRODUCTION

The transcription of protein coding genes is controlled by regulatory DNA elements, including both the core promoter and more distal enhancer elements (Lee and Young, 2000). The core promoter is a short DNA region surrounding the transcription start site (TSS), at which RNA polymerase II and general transcription factors are recruited. Enhancer elements act as platforms for recruiting both positive- and negative-acting transcription factors (TFs), and serve to integrate multiple signaling inputs in order to dictate the spatial and temporal control of transcription from the core promoter. As such, enhancer functions are critical for directing transcriptional output during cell differentiation and development, as well as coordinating transcriptional responses to environmental change (Ong and Corces, 2011). Despite their importance, only a small number of *bona fide* enhancers have been characterized in plants, and we lack a global view of their general distribution and action in plant genomes (Weber et al., 2016).

In large part, our limited knowledge of plant *cis*-regulatory elements arises from the unique difficulties in identifying these elements. While some enhancers exist near their target core promoter, others can be thousands of base pairs upstream or downstream, or even within the transcribed region of a gene body (Ong and Corces, 2011; Spitz and Furlong, 2012). Furthermore, enhancers generally do not display universal sequence conservation, aside from sharing of individual TF binding sites, which makes them very challenging to locate. By contrast, core promoters can be readily identified through mapping the 5' ends of transcripts (Morton et al., 2014; Mejia-Guerra et al., 2015). It was recently discovered that many enhancer elements in animal genomes could be identified with relatively high confidence based on a unique combination of flanking histone posttranslational modifications (PTMs), such as an enrichment for H3K27ac and H3K4me1. This characteristic histone PTM signature has led to the annotation of such elements in several animal models and specialized cell types (Heintzman et al., 2009; Bonn et al., 2012). However, the only currently known association between plant *cis*-regulatory elements and histone PTMs

appears to be a modest correlation with H3K27me3 (Zhang et al., 2012b; Zhu et al., 2015). Though encouraging, this mark is not unique to these elements, and cannot be used to identify enhancers on its own.

A long-known and general feature of sequence-specific DNA-binding proteins is their ability to displace nucleosomes upon DNA binding, leading to an increase in nuclease accessibility around the binding region (Gross and Garrard, 1988; Henikoff, 2008). In particular, DNase I treatment of nuclei coupled with high-throughput sequencing (DNase-seq) has been used to probe chromatin accessibility. This technology has served as an important tool in identifying regulatory elements throughout animal genomes (Thurman et al., 2012) and more recently in certain plant genomes (Zhang et al., 2012b; Zhang et al., 2012a; Pajoro et al., 2014; Sullivan et al., 2014). In addition, a differential micrococcal nuclease sensitivity assay has also been used to probe functional regions of the maize genome, demonstrating the versatility of this approach (Vera et al., 2014; Rodgers-Melnick et al., 2016).

DNase-seq has been used successfully to identify open chromatin regions in different tissues of both rice and *Arabidopsis* (Zhang et al., 2012a; Pajoro et al., 2014; Zhu et al., 2015). Over a dozen of the intergenic DNase-hypersensitive sites in *Arabidopsis* were tested and shown to act as enhancer elements by activating a minimal promoter-reporter cassette, demonstrating that chromatin accessibility is an important factor in enhancer identification (Zhu et al., 2015). Collectively, these DNase-seq studies show that the majority of open chromatin sites exist outside of genes in rice and *Arabidopsis*, that differences in open chromatin sites can be identified between tissues, and that a large proportion of intergenic open chromatin sites are in fact regulatory, at least in *Arabidopsis*. Another recent significant advance came from using DNase-seq to examine the changes in *Arabidopsis* chromatin accessibility and TF occupancy that occur during development and in response to abiotic stress (Sullivan et al., 2014). This work showed that TF-to-TF regulatory network connectivity appears to be similar between *Arabidopsis*, human, and *C. elegans*, and that such networks were extensively 'rewired' in response to stress. This study also showed that many genetic variants linked to complex traits were preferentially located in accessible chromatin regions, portending the potential for harnessing natural variation in regulatory DNA for plant breeding.

We are still left with many open questions regarding the general conservation of transcriptional

regulatory landscapes across plant genomes. For example, it remains unclear how many *cis*-regulatory elements generally exist in plant genomes, where they reside in relation to their target genes, and to what extent these features are conserved across plant genomes. Furthermore, it is not clear how the *cis*-regulatory elements within a single genome confer cell type-specific transcriptional activity – and thus cell type identity – during development. In the present study, we seek to build on previous work and to address some of these outstanding questions by analyzing chromatin accessibility across multiple, diverse plant species, and between two distinct cell types.

From a methodological perspective, the DNase-seq procedure is relatively labor-intensive and requires a large number of starting nuclei for DNaseI treatment, which can be a major drawback for conducting cell type-specific profiling investigations. More recently, the Assay for Transposase-Accessible Chromatin with sequencing (ATAC-seq) was developed as an alternative approach (Buenrostro et al., 2013). ATAC-seq employs treatment of isolated nuclei with an engineered transposase that simultaneously cleaves DNA and inserts sequencing adapters, such that cleaved fragments originating from open chromatin can be converted into a high-throughput sequencing library by Polymerase Chain Reaction (PCR). Sequencing of the resulting library provides readout highly similar to that of DNase-seq, but ATAC-seq requires far fewer nuclei (Buenrostro et al., 2015). The relatively simple procedure for ATAC-seq and its low nuclei input, combined with its recent application in *Arabidopsis* and rice (Wilkins et al., 2016; Bajic et al., 2017; Lu et al., 2017), has made it widely useful for assaying plant DNA regulatory regions. In this study, we first optimized ATAC-seq for use with crude nuclei and nuclei isolated by INTACT (Isolation of Nuclei TAgged in specific Cell Types) affinity purification (Deal and Henikoff, 2010). We then applied this method to INTACT-purified root tip nuclei from *Arabidopsis thaliana*, *Medicago truncatula*, *Solanum lycopersicum* (tomato), and *Oryza sativa* (rice), as well as the root hair and non-hair epidermal cell types of *Arabidopsis*. The use of diverse plant species of both dicot and monocot lineages allowed us to assay regulatory structure over a broad range of evolutionary distances. Additionally, analysis of the *Arabidopsis* root hair and non-hair cell types allowed us to identify distinctions in chromatin accessibility that occurred during the differentiation of developmentally linked cell types from a common progenitor stem cell.

In our cross-species comparisons, we discovered that the majority of open chromatin sites in all four species exist outside of transcribed regions. The open sites also tended to cluster within several kilobases upstream of the transcription start sites despite the large differences in intergenic space between the four genomes. When orthologous genes were compared across species, we found that the number and location of open chromatin regions were highly variable, suggesting that regulatory elements are not statically positioned relative to target genes over evolutionary timescales. However, we found evidence that particular gene sets remain under control by common TFs across these species. For instance, we discovered a set of four TFs that appear to be integral for root tip transcriptional regulation of common gene sets in all species. These include HY5 and MYB77, which were previously shown to impact root development in *Arabidopsis* (Oyama et al., 1997; Shin et al., 2007).

When comparing the two *Arabidopsis* root epidermal cell types, we found that their open chromatin profiles are qualitatively very similar. However, many quantitative differences between cell types were identified, and these regions often contained binding motifs for TFs that were more highly expressed in one cell type than the other. Further analysis of several such cell type-enriched TFs led to the discovery of a hair cell transcriptional regulatory module driven by ABI5 and MYB33. These factors appear to co-regulate a number of additional hair cell-enriched TFs, including MYB44 and MYB77, which in turn regulate many downstream TF genes as well as other genes impacting hair-cell fate, physiology, secondary metabolism, and stress responses.

Overall, our work suggests that the *cis*-regulatory structure of these four plant genomes is strikingly similar, and that TF-target gene modules are also generally conserved across species. Furthermore, early differential expression of high-level TFs between the *Arabidopsis* hair and non-hair cells appears to drive a TF cascade that at least partially explains distinctions between hair and non-hair cell transcriptomes. Our data also highlight the utility of comparative chromatin profiling approaches and will be widely useful for hypothesis generation and testing.

RESULTS AND DISCUSSION

Application of ATAC-seq in *Arabidopsis* root tips

The Assay for Transposase-Accessible Chromatin (ATAC-seq) method was introduced in 2013 and has since been widely adopted in many systems (Buenrostro et al., 2013; Mo et al., 2015; Scharer et al., 2016; Lu et al., 2017). This technique utilizes a hyperactive Tn5 transposase that is pre-loaded with sequencing adapters as a probe for chromatin accessibility. When purified nuclei are treated with the transposase complex, the enzyme freely enters the nuclei and cleaves accessible DNA, both around nucleosomes and at nucleosome-depleted regions arising from the binding of transcription factors (TFs) to DNA. Upon cleavage of DNA, the transposon integrates sequencing adapters, fragmenting the DNA sample in the process. Regions of higher accessibility will be cleaved by the transposase more frequently and generate more fragments – and ultimately more reads, once the sample is sequenced. Conversely, less accessible regions will have fewer fragments and reads. After PCR-amplification of the raw DNA fragments, paired-end sequencing of the ATAC-seq library can reveal nucleosome-depleted regions where TFs are bound.

In this study, we set out to apply ATAC-seq to multiple plant species as well as different cell types from a single species. As such, we first established procedures for using the method with *Arabidopsis*, starting with root tip nuclei affinity-purified by INTACT (Isolation of Nuclei TAgged in specific Cell Types). We also established a protocol to use nuclei purified by detergent lysis of organelles followed by sucrose sedimentation, with the goal of broadening the application of ATAC-seq to non-transgenic starting tissue. We began with an *Arabidopsis* INTACT transgenic line constitutively expressing both the nuclear envelope targeting fusion protein (NTF) and biotin ligase (BirA) transgenes. Co-expression of these transgenes results in all the nuclei in the plant becoming biotinylated, and thus amenable to purification with streptavidin beads (Deal and Henikoff, 2010; Sullivan et al., 2014). Transgenic INTACT plants were grown on vertically oriented nutrient agar plates to facilitate root growth, and total nuclei were isolated from the 1 cm root tip region. These nuclei were further purified either by treatment with 1% (v/v) Triton X-100 and sedimentation through a sucrose cushion (‘Crude’ purification) or affinity-purified using streptavidin-coated magnetic beads (INTACT purification). In both cases 50,000 nuclei from each

purification strategy were used as the input for ATAC-seq (Figure 3.1A). Overall, both Crude and INTACT-purified nuclei yielded very similar results (Figure 3.1B and C, Figure S3.1). One clear difference that emerged was the number of reads that map to organellar DNA between the nuclei preparation methods. While the total reads of Crude nuclei preparations mapped approximately 50% to organellar genomes and 50% to the nuclear genome, the total reads of INTACT-purified nuclei consistently mapped over 90% to the nuclear genome (Table 3.1). The issue of organellar genomes contaminating ATAC-seq reactions is a common one, resulting in a large percentage of organelle-derived reads that must be discarded before further analysis. This issue was also recently shown to be remedied by increasing the purity of nuclei prior to ATAC-seq by use of fluorescence-activated nuclei sorting (Lu et al., 2017). To compare between datasets for the Crude and INTACT preparation strategies, we analyzed the enrichment of ATAC-seq reads using Hotspot peak mapping software (John et al., 2011). Though designed for use with DNase-seq data, Hotspot can also be readily used with ATAC-seq data. The number of enriched regions found with this algorithm did not differ greatly between nuclei preparation types, nor did the SPOT score (a signal-specificity measurement representing the proportion of sequenced reads that fall into enriched regions) (Table 3.1). These results suggest that the datasets are generally comparable regardless of the nuclei purification method.

Visualization of the Crude- and INTACT-ATAC-seq datasets in a genome browser revealed that they were highly similar to one another and to DNase-seq data from whole root tissue (Figure 3.1B). Further evidence of similarity among these datasets was found by examining the normalized read count signal in all datasets (both ATAC-seq and DNase-seq) within the regions called as ‘enriched’ in the INTACT-ATAC-seq dataset. For this and all subsequent peak calling in this study, we used the *findpeaks* algorithm in the HOMER package (Heinz, Benner et al. 2010), which we found to be more versatile and user-friendly than Hotspot. Using this approach, we identified 23,288 enriched regions in our INTACT-ATAC-seq data. We refer to these peaks, or enriched regions, in the ATAC-seq data as transposase hypersensitive sites (THSs). We examined the signal at these regions in the whole root DNase-seq dataset and both Crude- and INTACT-ATAC-seq datasets using heatmaps and average plots. These analyses showed that THSs detected

in INTACT-ATAC-seq tended to be enriched in both Crude-ATAC-seq and DNase-seq signal (Figure 3.1C). In addition, the majority of enriched regions (19,516 of 23,288) were found to overlap between the root-tip INTACT-ATAC-seq and the whole-root DNase-seq data (Figure 3.1D) and the signal intensity over DNase-seq or ATAC-seq enriched regions was highly correlated between the datasets (Figure S3.1).

To examine the distribution of hypersensitive sites among datasets, we identified enriched regions in both types of ATAC-seq datasets and the DNase-seq dataset, and then mapped these regions to genomic features. We found that the distribution of open chromatin regions relative to gene features was nearly indistinguishable among the datasets (Figure 3.1E). In all cases, the majority of THSs (~75%) were outside of transcribed regions, with most falling within 2 kb upstream of a transcription start site (TSS) and within 1 kb downstream of a transcript termination site (TTS).

Overall, these results show that ATAC-seq can be performed effectively using either Crude or INTACT-purified nuclei, and that the data in either case are highly comparable to that of DNase-seq. While the use of crudely purified nuclei should be widely useful for assaying any tissue of choice without a need for transgenics, it comes with the drawback that ~50% of the obtained reads will be from organellar DNA. The use of INTACT-purified nuclei greatly increases the cost efficiency of the procedure and can also provide access to specific cell types, but requires pre-established transgenic lines.

Comparison of root tip open chromatin profiles among four species

Having established an efficient procedure for using ATAC-seq on INTACT affinity-purified nuclei, we used this tool to compare the open chromatin landscapes among four different plant species. In addition to the *Arabidopsis* INTACT line described above, we also generated constitutive INTACT transgenic plants of *Medicago truncatula* (*Medicago*), *Oryza sativa* (rice), and *Solanum lycopersicum* (tomato). Seedlings of each species were grown on vertically oriented nutrient plates for one week after radicle emergence, and nuclei from the 1 cm root tip regions of each seedling were isolated and purified with streptavidin beads. ATAC-seq was performed in at least two biological replicates for each species, starting with 50,000 purified nuclei in each case. Visualization of the mapped reads across each genome showed notable consistencies

in the data for all four species. In all cases, the reads localize to discrete peaks that are distributed across the genome, as expected (Figure 3.2A). Examination of a syntenic region found in all four genomes suggested at least some degree of consistency in the patterns of transposase accessibility around orthologous genes (Figure 3.2A).

To specifically identify regions of each genome that were enriched in ATAC-seq signal (THSs), we used the HOMER *findpeaks* function on each biological replicate experiment. For further analysis, we retained only THS regions that were found in at least two biological replicates of ATAC-seq in each species. These reproducible THSs were then mapped to genomic features in each species in order to examine their distributions. As seen previously for *Arabidopsis*, the majority of THSs (~70-80%) were found outside of transcribed regions in all four species (Figure 3.2B). For this analysis, we classified these extragenic THSs (THSs found anywhere outside of transcribed regions) as proximal upstream (< 2 kb upstream of the transcription start site, or TSS), proximal downstream (< 1 kb downstream of the transcript termination site, or TTS) or intergenic (> 2 kb upstream from a TSS or > 1 kb downstream from a TTS). The proportion of THSs in the proximal upstream and intergenic regions varied greatly with genome size, and thus the amount of intergenic space in the genome. For example, a full 52% of THSs in *Arabidopsis* – the organism with the smallest genome (~120 Mb) and highest gene density of the four species – were in the proximal upstream region. This percentage drops as genome size and intergenic space increase, with 37% of the THSs in the proximal upstream region in the rice genome (~400 Mb), 30% in the *Medicago* genome (~480 Mb), and a mere 11% in the tomato genome (~820 Mb). The percentage of total THSs in the proximal downstream region followed a similar pattern, marking 17% of the THSs in *Arabidopsis*, 12% in rice and *Medicago*, and 6% in tomato. Finally, the proportion of THSs classified as intergenic followed the inverse trend as expected, with 12% of the THSs in intergenic regions for *Arabidopsis*, 30% for rice and *Medicago*, and 50% for tomato (Figure 3.2B). Thus, while the overall proportion of extragenic THSs is similar among species, the distance of these sites from genes tends to increase with genome size, which is roughly proportional to the average distance between genes.

Since the majority of THSs were found upstream of the nearest gene for each species, we next classified the regions based on their distance from the nearest TSS.\} We binned THSs in each genome into twelve distance categories, starting with those > 10 kb upstream of the TSS, then into eleven bins of 999 bp moving in toward the TSS, and finally a TSS-proximal bin of 100-0 bp upstream of the TSS (Figure 3.2C). Starting with this TSS-proximal bin, we find that ~17% of the upstream THSs in *Arabidopsis*, *Medicago*, and rice are within 100 bp of the TSS, whereas 2.7% of the upstream THSs in tomato are within 100 bp of the TSS. Moving away from the TSS, we find that 91% of the total upstream THSs fall within 2.9 kb of the TSS in *Arabidopsis*, while this number decreases with genome size, with 84% for rice, 73% for *Medicago*, and 65% for tomato. In the distance bin spanning 9.9 kb to 3 kb upstream, we find 7% of the total upstream THSs in *Arabidopsis*, 15% in rice, 23% in *Medicago*, and 32% in tomato. Finally, the THSs that are more than 10 kb away from the TSS accounts for 0.8% of the total upstream THSs in *Arabidopsis*, 0.9% in rice, 2.3% in *Medicago*, and 3.3% in tomato. Overall, it is clear that in all species the majority of THSs are within 3 kb upstream of a TSS, suggesting that most *cis*-regulatory elements in these genomes are likely to be proximal to the core promoter. In the species with the largest genomes and intergenic distances (*Medicago* and tomato), THSs tend to be spread over a somewhat wider range upstream of the TSS. However, even in these cases, only a few hundred THSs in total are more than 10 kb away from the nearest gene. It is worth noting that the distribution of THSs in *Medicago* is more similar to that of tomato than rice, despite the genome size being more similar to rice. This suggests that THSs tend to be further away from TSSs in *Medicago* than would be expected based on genome size alone.

As most THSs fall near genes, we next investigated from the opposite perspective – for any given gene, how many THSs were associated with it? In this regard, we find that the *Arabidopsis*, *Medicago*, and rice genomes are highly similar (Figure 3.2D). In all three genomes, of the subset of genes that have *any* upstream THSs, ~70% of these genes have a single site, ~20% have two sites, 5-7% have three sites, and 2-3% have four or more THSs. By contrast, the tomato genome has a different trend. Of the subset of tomato genes with *any* upstream THS, only 27% of the genes have a single site, and this proportion gradually

decreases with increasing THS number, with 2.7% of the tomato genes in this subset having 10 or more THSs.

Overall, we have found that THSs have similar size and genomic distribution characteristics across all four species (Table S3.1). The majority of THSs in all species are found outside of genes, mainly upstream of the TSS, and these sites tend to cluster within 3 kb of the TSS. Furthermore, most genes with an upstream THS in *Arabidopsis*, *Medicago*, and rice have only 1-2 THSs, whereas tomato genes tend to have a larger number of upstream THSs. Whether this increase in upstream THSs in tomato is reflective of an increase in the number of regulatory elements per gene based on clade-specific alterations in gene regulation, DNA copy number changes, or simply the greater abundance of transposons and other repeat elements is not entirely clear. Compared to the other species, tomato THSs are much more abundant and tend to be smaller in size than those of the other species, and the tomato ATAC-seq data generally appear to have a lower signal-to-noise ratio (Table S3.8, Figure 3.2A). While it is unclear why the data from tomato are distinct in these ways, it is clear that tomato THSs occupy mostly genic regions of the genome, as expected, and are highly reproducible between biological replicate experiments (Figure S3.2).

Collectively, these results suggest that there is a relatively small number of regulatory elements per gene in plants. These elements tend to be focused near the promoter rather than at more distal sites as has been observed in animal, particularly mammalian, genomes (Stadhouders et al., 2012). The assumptions implicit in this argument are that open chromatin sites near a TSS reflect regulatory elements that regulate that TSS and not a more distant one, and that upstream elements contribute the majority of regulatory effects. These assumptions appear to be generally validated by many reporter assays showing that an upstream fragment of several kilobases is frequently sufficient to recapitulate native transcription patterns (Medford et al., 1991; Masucci et al., 1996; Ruzicka et al., 2007; Tittarelli et al., 2009; Li et al., 2012), as well as our observation that upstream THSs are the most abundant class of open chromatin sites.

Open chromatin features are not directly conserved among orthologous genes

Given that many of the properties of open chromatin regions were shared among *Arabidopsis*, *Medicago*, rice, and tomato, we next asked whether the numbers and locations of THSs – and thus putative regulatory elements – were conserved among orthologous genes across species. For these analyses, we identified 373 syntenic orthologs (Table S3.2) that were found in all four genomes and asked whether members of each ortholog set harbored a similar number of open chromatin regions across the species. Again, using root tip THSs present in at least two biological replicates for each species, we counted the number of THSs within 5 kb upstream of the TSS for each ortholog in each species. We then examined these data for similarities and differences in upstream THS number (Figure 3.3A). While no clear trend of strong conservation in the number of upstream THSs emerged from this analysis, there was a small subset of orthologs that did have upstream THSs in similar numbers across species. However, this was a very small proportion of the total. As seen in earlier analyses, tomato genes tended to have a larger number of upstream THSs compared to the other species, and most of the 373 orthologs in tomato did have at least one upstream THS. This was not the case in the other three species, where many of the orthologs had no detectable upstream THSs within 5 kb of the TSS. Among the four species, *Arabidopsis* and *Medicago* showed the greatest similarity in upstream THS number, but even in this case the similarity was minimal despite the relatively closer phylogenetic relationship between these two organisms.

We next examined the distribution of open chromatin regions across the upstream regions of these 373 orthologous genes relative to their expression level in *Arabidopsis*, reasoning that there could be patterns of open chromatin similarity based on THS positions, rather than numbers. For this analysis, we examined the normalized ATAC-seq signal across the upstream region of all 373 orthologous genes, from -5000 bp to +100 bp relative to the TSS of each gene (Figure 3.3B). Orthologs were then ranked within the heatmap based on the transcript level of each *Arabidopsis* ortholog in the root tip (Li, Yamada et al. 2016), from highest to lowest expression. For each *Arabidopsis* ortholog we also included the upstream THS number to ascertain how this feature might correlate with transcript level for *Arabidopsis*. While there was some consistency among species in that open chromatin often overlapped with the TSS, we did not observe any clear pattern in transposase hypersensitivity within the upstream regions of these orthologs. K-means

clustering of the heatmaps similarly did not reveal evidence for conservation of open chromatin patterns among orthologs (Figure S3.3A). An important caveat to this analysis is that many of these syntenic orthologs may not be functional homologs, or 'expressologs' (Patel et al., 2012), due to subfunctionalization within gene families. As such, we identified a smaller group (52) of expressologs on which to perform a similar test (Table S3.3). While these expressolog genes have both maximally high protein level similarity and expression pattern similarity, including expression in the root, there was also no clear correspondence in upstream THS number among them (Figure S3.3B).

There does not appear to be strong conservation in the number and location of open chromatin sites at orthologous genes across species. Assuming that these genes are still under control of common TFs, this suggests that regulatory elements could be free to migrate, and perhaps split or fuse, while retaining the regulatory parameters of the target gene in question.

One interesting finding from these analyses was that the pattern of upstream THS number does not correlate with expression level, at least for *Arabidopsis* (Figure 3.3B). Thus, THSs must not simply represent activating events upstream of the TSS but may also represent binding of repressive factors. Further, we found no correlation between upstream THS number and expression entropy among all genes in the *Arabidopsis* genome, suggesting a more complex relationship between regulatory element distribution and target gene transcription (Figure S3.3C).

Evidence for co-regulation of common gene sets by multiple TFs across species

While there does not appear to be a consistent pattern in the number or placement of open chromatin regions around orthologs or expressologs, we wanted to examine whether it would be possible to find common regulators of specific gene sets among species using a deeper level of analysis. To do this, we first searched for common TF motifs in root tip THSs across the four species. Using the THSs that were found in at least two replicates for each species, we employed the MEME-ChIP motif analysis package (Machanick and Bailey, 2011; Ma et al., 2014) to identify overrepresented motifs of known TFs. We discovered 30 motifs that were both overrepresented and common among all species (Table S3.4). We

narrowed our list of candidate TFs by considering a variety of factors, including the expression of each TF in the root tip, any known mutant root phenotypes involving those TFs, and whether genome-wide binding information was available for each candidate in *Arabidopsis*. Ultimately, we selected 4 TFs for further analysis: ELONGATED HYPOCOTYL 5 (HY5), ABSCISIC ACID RESPONSIVE ELEMENTS-BINDING FACTOR 3 (ABF3), C-REPEAT/DRE BINDING FACTOR 2 (CBF2), and MYB DOMAIN PROTEIN 77 (MYB77). It is worth noting that among these factors, both HY5 and MYB77 had been previously implicated in root development (Oyama et al., 1997; Zhao et al., 2014). Like HY5 and MYB77, CBF2 and ABF3 have been implicated in stress responses as well as abscisic acid (ABA) signaling (Kang et al., 2002; Knight et al., 2004). Furthermore, overexpression of ABF3 leads to increased tolerance to multiple abiotic stresses in *Arabidopsis*, rice, cotton, and alfalfa (Oh et al., 2005; Abdeen et al., 2010; Wang et al., 2016; Kerr et al., 2017). Given this evidence, we decided to focus on these factors for further study.

We first sought to define the target genes for each of these four TFs in *Arabidopsis* by combining our chromatin accessibility data with published genome-wide binding data for each factor in *Arabidopsis* (Table 3.2). Because an accessible chromatin region (a THS) represents the displacement of nucleosomes by a DNA-binding protein, we reasoned that our THS profiles for a given tissue would represent virtually all possible protein binding sites in the epigenomes of root tip cells. Similarly, by using *in vitro* genomic binding data (DAP-seq) (O'Malley et al., 2016b) or ChIP-seq data from a highly heterogeneous tissue, we could identify the spectrum of possible binding sites for that TF, such that the intersection of these datasets would represent the binding sites for that TF in the sample of interest. While there are caveats to this approach, we reasoned that it was more likely to generate false negatives than false positives and would give us a set of high confidence target genes to analyze for each TF. In this regard, ChIP-seq data may be more robust because they represent *in vivo* binding, while DAP-seq is an *in vitro* assay and may not capture binding sites that depend on chromatin properties or interactions with other TFs. On the other hand, ChIP-seq data are inherently limited by the cell types present in the sample used.

We first tested this approach in *Arabidopsis* with each of the four TFs of interest. Using THSs from the *Arabidopsis* root tip that were found in at least two biological replicates, we used the motif-identification

tool FIMO (Grant et al., 2011) to identify THSs that contained a significant occurrence of the TF motif of interest. The THSs that contained a significant motif match were considered *predicted binding sites*. We then identified predicted binding sites that also overlapped with a known binding site for that TF (a DAP-seq or ChIP-seq peak), and these were considered *high confidence binding sites* for that TF in the root tip (Figure S3.4). The predicted binding sites (motif-containing THSs) were themselves very good predictors of the true binding sites for these four TFs (Table 3.2). For example, of the 1,316 *Arabidopsis* root tip THSs with an occurrence of the ABF3 motif (Mathelier, Zhao et al. 2014), 1,279 (97%) overlapped with an ABF3 ChIP-seq peak from whole 2-day-old seedlings (Song et al., 2016). Similarly, 89% of predicted CBF2 binding sites (Weirauch, Yang et al. 2014) overlapped with a CBF2 DAP-seq peak (O'Malley et al., 2016a), 74% of predicted MYB77 binding sites (Weirauch, Yang et al. 2014) overlapped with a MYB77 DAP-seq peak (O'Malley et al., 2016a), and 61% of predicted HY5 binding sites (Mathelier, Zhao et al. 2014) overlapped with a HY5 DAP-seq peak (O'Malley et al., 2016a). In each case, the high confidence binding sites (motif-containing THSs that overlap with a ChIP- or DAP-seq peak) were assigned to their nearest TSS in order to identify the putative target genes for each TF (Figure S3.4).

With these lists of target genes for each TF in the *Arabidopsis* root tip, we looked for gene sets that were regulated by more than one factor, as means of identifying co-regulatory associations between these four TFs. We found extensive co-targeting among these four TFs, with gene sets being targeted by one, two, three, or all four of these TFs to a degree that was far higher than what would be expected by chance (Figure 3.3C). For example, of the 1,271 ABF3 target genes, 297 (23%) are also targeted by HY5 (hypergeometric $p = 2.1 \times 10^{-56}$). Among these 297 genes, 46 are targeted by ABF3, HY5, and CBF2, and seven are targeted by all four TFs. We also asked where the binding sites driving this pattern were located relative to the target genes. To do this we considered only binding sites within the 5 kb upstream region of a TSS, and repeated the target gene assignment and analysis of target gene overlaps between TFs. This subsetting reduced the total number of target genes for each factor by ~20%, but did not substantially alter the percentages of target gene overlap among the four TFs (Figure S3.5A). These results collectively suggest that these four TFs have important roles in root tip gene regulation both individually and in

combination, and that the majority of their binding sites (~80%) fall within the 5 kb region upstream of the TSS for target genes. In addition, we find that the binding sites for multiple TFs often occur in the same THS (Figure S3.5B).

We next sought to examine the target genes and proportions of target gene overlaps between the four species to address the conservation of co-regulatory relationships among these four TFs. Given that no TF binding data is available for the other three species and knowing that the majority of our predicted binding sites in *Arabidopsis* corresponded to known binding sites (Table 3.2; 61-97%), we opted to also use the predicted binding sites for each of the four TFs in *Medicago*, tomato, and rice, with the knowledge that these sets may contain some false positives. For these analyses we used the *Arabidopsis* TF motifs – since these have not been directly defined for the other species – with the caveat that the DNA binding specificity of these factors may not be identical among species.

We again used FIMO to identify significant occurrences of each TF motif within the root tip THSs found in at least two biological replicates for each of our four species. We then mapped the predicted binding sites of each TF to the nearest TSS to define target genes for each TF in each species (Table S3.5). We then analyzed the overlap of TFs at target genes in each species using 4-way Venn diagrams, similar to Figure 3.3C. To compare regulatory associations across species, we considered each of the 15 categories in every species-specific 4-way Venn diagram as a regulatory category. For example, one regulatory category consists of the genes targeted only by ABF3 alone, another would be those targeted only by HY5 and ABF3 at the exclusion of the other two TFs, and so on. For each regulatory category in each species, we calculated the percentage of the total target genes in that category (number of genes in the regulatory category/total number of genes targeted by *any* of the 4 TFs), and then compared these percentages between species (Figure 3.3D). We found remarkably consistent proportions of the target genes in nearly all regulatory categories across all four species. However, notable deviations from this consistency among species were seen in the proportion of rice genes targeted by MYB77 alone and rice genes targeted CBF2 and HY5 together. In most cases, the proportions of target genes in different regulatory categories were most similar between *Arabidopsis* and *Medicago*, and these were generally more similar to tomato than to rice, consistent

with the evolutionary distances between the species (Vanneste et al., 2014). Commonly overrepresented Gene Ontology (GO) terms among gene sets in particular regulatory categories across species further support the notion of regulatory conservation (Figure S3.5C), although these analyses are limited by the depth of GO annotation in some of these species.

These findings suggest that while neither syntenic orthologous gene sets nor expressolog gene sets tend to share open chromatin patterns, the genes under control of specific TFs or specific combinations of TFs appear to be relatively stable over evolutionary time, at least for the four TFs we examined. One simple explanation for this phenomenon is that the locations of transcriptional regulatory elements are somewhat malleable over time as long as proper transcriptional control is maintained. In this model, these elements would be free to relocate in either direction, and potentially even merge or split. This would maintain proper control over the target gene, but give each ortholog or expressolog a unique chromatin accessibility profile depending on the exact morphology and distribution of the functionally conserved regulatory elements. This idea of modularity is consistent with previous observations that the *Drosophila* even-skipped stripe 2 enhancer can be rearranged and still retain functionality (Ludwig et al., 2000; Ludwig et al., 2005).

The results also shed light on the interconnectedness of specific TFs in root tip cells and indicate durability of these co-regulatory relationships over time. They also generate readily testable hypotheses regarding the how HY5, ABF3, MYB77, and CBF2 operate during root development. For example, given that HY5 appears to regulate over 1,000 genes in the *Arabidopsis* root tip (Figure 3.3C), and that hundreds of these are annotated with GO terms including *biological regulation* and *response to stimulus*, we predict that *hy5* mutants would have defects in root tip morphology and growth. Indeed, HY5 was previously shown to be involved in the regulation of lateral root growth initiation and gravitropism (Oyama, Shimura et al. 1997), and we observe that the primary root tips in *hy5* mutants also frequently show a bulging and malformed appearance, as well as severe gravitropism defects (Figure S3.6).

Commonalities and distinctions in the open chromatin landscapes of *Arabidopsis* root epidermal cell types

Having examined questions of regulatory conservation between species, we then explored regulatory elements and TFs relationships between cell types within a single species. In this case, we chose to focus on the root epidermal hair and non-hair cell types in *Arabidopsis*. Since these two cell types are derived from a common progenitor, they are prime candidates to offer insight into the epigenomic alterations that occur during – and likely drive – cell differentiation. Specifically, we investigated to what extent the open chromatin landscapes would differ between cell types and whether differences in THSs could pinpoint the sites of differential transcriptional regulation. Furthermore, we wanted to understand whether we could use this information to examine the TF-to-TF regulatory connections that underlie the transcriptomic and physiological differences between these cell types.

We used two previously described INTACT transgenic lines as starting material for these experiments: one having biotin-labeled nuclei exclusively in the root hair (H) cells, and another with labeled nuclei only in the root epidermal non-hair (NH) cells (Deal and Henikoff, 2010). Nuclei were purified from each fully differentiated cell type by INTACT, and 50,000 nuclei of each type were subjected to ATAC-seq. Visualization of these cell type-specific datasets in a genome browser, along with the *Arabidopsis* whole 1 cm root tip ATAC-seq data, showed a high overall degree of similarity among the three datasets (Figure 3.4A). Comparison of the ATAC-seq signal intensity at common THS regions genome-wide revealed that these two cell types have open chromatin patterns that are highly similar to one another, but distinct from that of the whole root tip (Figure S3.7).

To identify regions of differential accessibility between the cell types and the whole root tip, we considered THS regions that were found in at least two biological replicates of each cell type or tissue. The total number of these reproducible THSs was 32,942 in the whole root tip, 35,552 for the H cells, and 28,912 for the NH cells. The majority of these sites (18,742) were common (overlapping) in all three sample types (Figure 3.4B) and thus likely represent regulatory sites that are utilized in multiple *Arabidopsis* root cell types. We also found 6,562 THSs that were common to both root epidermal cell types but were not found in the whole root tip, suggesting that these may represent epidermal-specific regulatory elements. In a search for unique THSs in each of the three sample types (those not overlapping with a THS in any other

sample), we found 10,455 THSs that were unique to the whole root tip, 7,537 unique to the H cells, and 2,574 that were unique to the NH cells. We refer to these regions as differential THSs (dTHSs). The dTHSs identified only in the H or NH cell type were of further interest because they may represent regulatory elements that drive the transcriptomic differences between these two epidermal cell types.

To examine the extent of chromatin accessibility differences at these dTHSs, we visualized the accessibility signals from each cell type at both H cell dTHSs and NH cell dTHSs. First, using the 7,537 regions identified as H cell dTHSs, we used heatmaps and average plots to examine the normalized ATAC-seq read count across these regions in each cell type (Figure 3.4C, left panel). We then repeated this analysis using the 2,574 NH cell dTHSs (Figure 3.4C, right panel). In each case, it was clear that the regions we identified as dTHSs showed significant differences in chromatin accessibility between the two cell types. However, the differences in chromatin accessibility between cell types were quantitative (varying intensity) rather than qualitative (all-or-nothing). This indicates that, at large, the dTHSs represent sites that are highly accessible in one cell type and less so in the other, rather than being strictly present in one and absent in the other. Therefore, we refer to these sites from this point on as cell type-enriched dTHSs to convey the notion of quantitative differences between cell types.

To identify the genes that might be impacted by cell type-enriched dTHSs, we mapped each dTHS to its nearest TSS and considered that to be the target gene. We found that the 7,537 H-enriched dTHSs mapped to 6,008 genes, while the 2,574 NH-enriched dTHSs mapped to 2,295 genes. Thus, the majority of genes that are associated with a dTHS are only associated with one such site. This is consistent with our previous findings that most *Arabidopsis* genes are associated with a single upstream THS (Figure 3.2D).

We then asked how the set of genes associated with dTHSs overlapped with those whose transcripts that show differential abundance between the two cell types. Using data from a recent comprehensive RNA-seq analysis of flow sorted *Arabidopsis* root cell types (Li et al., 2016a), we identified sets of transcripts that were more highly expressed in H versus NH cell types. To be considered a *cell type-enriched gene*, we required a gene to have a transcript level with two-fold or greater difference in abundance between H and NH cell types, as well as at least five reads per kilobase per million mapped reads (RPKM) in the cell type

with a higher transcript level. Using this relatively conservative approach, we derived a list of 3,282 H cell-enriched genes and 2,731 NH cell-enriched genes. We then asked whether the genes associated with cell type-enriched dTHSs were also cell type-enriched genes (Figure 3.4D). Of the 3,282 H cell-enriched genes, 743 were associated with a H cell-enriched dTHS, 258 were associated with a NH cell-enriched dTHS, and 108 genes were associated with a dTHS in both cell types. Among the 2,731 NH cell-enriched genes, 156 were associated with a NH cell-enriched dTHS, 516 were associated with a H cell-enriched dTHS, and 52 genes showed dTHSs in both cell types. These results suggest that cell type-enriched expression of a gene is frequently associated with a dTHS in the cell type where the gene is highly expressed, but is also often associated with a dTHS in the cell type where that gene is repressed. This highlights the importance of transcriptional activating events in the former case and repressive events in the latter. Interestingly, for a smaller set of cell type-enriched genes we observed dTHSs at a given gene in both cell types, indicating regulatory activity at the gene in both cell types.

We next asked what proportion of the transcriptome differences between H and NH cells might be explained based on differential chromatin accessibility. Of the 3,282 H cell-enriched genes, 1,109 have a dTHS in one or both of the cell types, and among the 2,731 NH cell-specific genes, 724 have a dTHS in one or both cell types. Assuming that each dTHS represents a regulatory event contributing to the differential expression of its identified target gene, we could explain differential expression of 33% of the H cell-enriched genes and 27% of the NH cell-enriched genes. The remaining ~70% of the identified cell type-enriched genes without clear chromatin accessibility differences may be explained in numerous ways. These genes may not require a change in chromatin accessibility, changes in chromatin accessibility may fall below our limit of detection, or these transcripts may be primarily regulated at the post-transcriptional level rather than at the chromatin-accessibility level that we measured.

Another key question relates to the significance of the cell-type-enriched dTHSs that do not map to differentially expressed genes. These could be explained by an inability to detect all differentially expressed genes, perhaps simply due to the stringency of our definition of cell type-enriched genes. An important biological possibility to consider is that many of these regulatory regions do not in fact regulate

the closest gene, but rather act over a distance such that they are orphaned from their true target genes in our analysis. Another possibility is that many of the differential protein binding events represented by these dTHSs are unrelated to transcriptional regulation.

Overall, the accessible chromatin landscapes of the root epidermal H and NH cells appear to be nearly identical in a qualitative sense, but differ significantly at several thousand sites in each cell type. The reasons for the quantitative, rather than all-or-nothing, nature of this phenomenon are not entirely clear. Are the accessibility differences between cell types reflective of unique protein assemblages at the same element in different cell types, or do they instead reflect differences in abundance of the same proteins at an element in different cell types? While these questions certainly warrant further investigation and experimentation, we can gain further insight into the regulatory differences between cell types through deeper examination of the differentially accessible chromatin regions in each.

TF motifs in cell type-specific THSs identify regulators and their target genes

As a means of identifying specific transcription factors (TFs) that might be important in specifying the H and NH cell fates, we sought to identify overrepresented motifs in the differentially accessible regions of each cell type. We used each set of cell type-enriched dTHSs as input for MEME-ChIP analyses (Machanick and Bailey, 2011) and examined the resulting lists of overrepresented motifs. We initially found 219 motifs that were significantly overrepresented relative to genomic background only in H cell-enriched dTHSs and 12 that were significantly overrepresented only in NH cell-enriched dTHSs (Table S3.6). In order to narrow our list of candidate TFs to pursue, we vetted these lists of potential cell type-enriched TFs by considering their transcript levels in each cell type as well as the availability of genome-wide binding data. Based on the available data, we narrowed our search to five transcription factors of interest: four H cell-enriched TF genes (MYB33, ABI5, NAC083, and At5g04390) and one NH-enriched TF gene (WRKY27) (Table 3.3).

We next attempted to directly identify the binding sites for each TF by differential ATAC-seq footprinting between the cell types. The logic behind this approach is the same as that for DNase-seq

footprinting – that the regions around a TF binding site are hypersensitive to the nuclease or transposase due to nucleosome displacement, but the sites of physical contact between the TF and DNA will be protected from transposon insertion/cutting, and thus leave behind a characteristic "footprint" of reduced accessibility on a background of high accessibility (Hesselberth et al., 2009; Vierstra and Stamatoyannopoulos, 2016). We reasoned that we could identify binding sites for each of these cell type-enriched TFs by comparing the footprint signal at each predicted binding site (a motif occurrence within a THS) between H and NH cells.

For this analysis, we examined the transposase integration patterns around the motifs of each TF in both cell types as well as in purified genomic DNA subjected to ATAC-seq, to control for transposase sequence bias. It was recently reported in *Arabidopsis* that many TF motifs exhibit conspicuous transposase integration bias on naked DNA (Lu et al., 2017), and our results were in line with these findings for all five TFs of interest here (Figure S3.8). While we observed footprint-like patterns in the motif-containing THSs in our ATAC-seq data, these patterns in each case were also evident on purified genomic DNA. As such, it was not possible to distinguish true binding sites from these data, as any footprint signal arising from TF binding was already obscured by the transposase integration bias. For unknown reasons, many TF motif DNA sequences seem to inherently evoke hyper- and/or hypo-integration by the transposase, and this automatically obscures any potentially informative footprint signal that could be obtained by integration during ATAC-seq on nuclei. Similar technical concerns have also been raised for DNaseI footprinting (Sung et al., 2016). These results suggest that the ATAC-seq footprinting approach may be useful for certain TFs, but these will likely need to be examined on a case-by-case basis. Given this issue and the resulting lack of evidence for footprints of our TFs of interest, we decided to take the approach of defining TF target sites as we did for our studies of root tip TFs.

As described earlier, we defined high confidence binding sites for the 5 TFs of interest as TF motif-containing THSs in the cell type of interest (predicted binding sites) that *also* overlapped with an enriched region for the TF in publicly available DAP-seq data (O'Malley et al., 2016a) or ChIP-seq data (Figure S3.4). Assigning these high confidence binding sites to their nearest TSS allowed us to define thousands of

target genes for these factors in the root epidermal cell types (Table 3.3 and Table S3.7). Compared to our analysis of root tip TFs, our capability to predict target sites based on motif occurrences in THSs was much reduced for the four H cell-enriched and one NH cell-enriched TFs examined here. For further analyses, we decide to focus on three of the TFs that were more highly expressed in the H cell type and had the largest number of high confidence target genes: ABI5, MYB33, and NAC083.

We first asked how many of the high confidence target genes for these TFs were also preferentially expressed in one cell type or the other. We found that for all three TFs, a large percentage of the total target genes are H cell-enriched in their expression (17-21%), while many others are NH cell-enriched (6-9%) (Figure 3.5A). These results are intriguing as they suggest that the activities of these TFs may be generally context-dependent. At the same time however, the majority of the target genes for each TF were not more highly expressed in one cell type compared to the other.

Each of these H cell-enriched TFs could activate other H cell-enriched genes, but what are their functions at regulatory elements near genes that are expressed at low levels in the H cell and high levels in the NH cell? One possibility is that these factors are activators of transcription in the context of H cell-enriched genes but act as repressors or are neutral toward the target genes that are NH cell-enriched in their expression. This may reflect context-dependency in the sense that the effect on transcription of a target gene may depend on the local milieu of other factors.

We next examined whether ABI5, MYB33, and NAC083 target any of the same genes. Similar to the root tip TFs examined previously, we found that these three TFs also appear to have extensive co-regulatory relationships (Figure 3.5B). For example, 207 target genes were shared between ABI5 and NAC083, 238 were shared between ABI5 and MYB33, and 50 target genes were shared by all 3 factors. We further analyzed the genes that were co-targeted by ABI5 and MYB33, finding that 57 of the co-targeted genes were H-cell enriched. As such, we performed Gene Ontology (GO) analysis on the H cell-enriched targets as well as the full set of target genes to gain insight into the functions of this co-regulatory relationship (Figure 3.5C). Many of the ABI5/MYB33 target genes were annotated as being involved in responses to ABA as well as water, salt, and cold stress. This is consistent with the known roles of these

proteins in ABA signaling (Finkelstein and Lynch, 2000; Reyes and Chua, 2007). Interesting, seven of the 57 ABI5/MYB33 target genes that were H cell-enriched were also annotated with the term *regulation of transcription*, suggesting that ABI5 and MYB33 may be at the apex of a transcriptional regulatory cascade in the H cell type.

Identification of a new regulatory module in the root hair cell type

Based on our findings that ABI5 and MYB33 co-target seven H cell-enriched TFs, we decided to investigate this potential pathway further. Among the seven TFs putatively co-regulated by ABI5 and MYB33 and having H cell-enriched transcript expression were DEAR5, ERF11, At3g49930, SCL8, NAC087, and two additional MYB factors: MYB44 and MYB77. Aside from MYB77, none of these TFs had been previously reported to produce root-specific phenotypes when mutated. MYB77 was previously shown to interact with Auxin Response Factors (ARFs) (Shin et al., 2007) and to be involved in lateral root development through promotion of auxin-responsive gene expression (Shin et al., 2007). Interestingly, the ABA receptor, PYL8, was shown to physically interact with both MYB77 and MYB44, and to promote auxin-responsive transcription by MYB77 (Zhao et al., 2014). MYB44 has also been implicated in ABA signaling through direct interaction with an additional ABA receptor, PYL9 (Li et al., 2014), as well as repression of jasmonic acid (JA)-responsive transcription (Jung et al., 2010). These factors have additionally been implicated in salicylic acid (SA) and ethylene signaling (Yanhui et al., 2006; Shim et al., 2013). Given that MYB44 and MYB77 are paralogs (Dubos et al., 2010) that appear to integrate multiple hormone response pathways in a partly redundant manner (Jaradat et al., 2013), we decided to identify high confidence target genes (Figure S3.4) for each of them for further study.

We again defined high confidence binding sites as THSs in H cells that contain a significant motif occurrence for the factor and also overlap with a DAP-seq or ChIP-seq enriched region for that factor. Using this approach, we found that MYB44 and MYB77 each target over 1,000 genes individually and co-target 483 genes (Figure 3.6A). In addition, MYB44 and MYB77 appear to regulate one another, while MYB77 also appears to target itself. This feature of self-reinforcing co-regulation could serve as an

amplifying and sustaining mechanism to maintain the activity of this module once activated by ABI5, MYB33, and potentially other upstream factors.

To gain a deeper understanding of the impact of MYB44 and MYB77 on downstream processes, we performed Gene Ontology (GO) analysis of the target genes for each factor. First considering all target genes, regardless of their expression in the H cell type, we found a variety of overrepresented GO terms for each that were consistent with the known roles of these factors in hormone signaling (Figure 3.6B). For example, both factors targeted a large number of genes annotated with the terms *response to ABA stimulus*, *response to ethylene stimulus*, and *response to SA stimulus*. Additionally, MYB44 alone targeted many genes with the annotation *response to JA stimulus*, consistent with its previously reported role as a negative regulator of JA signaling (Jung et al., 2010). Interestingly, the largest overrepresented gene functional category for both factors was *transcription factor activity* (102 genes for MYB77 and 183 genes for MYB44). This indeed further suggests that these factors initiate a cascade of transcriptional effects. The next-largest overrepresented term was *plasmodesma*, indicating that production and/or regulation of cell-cell connecting structures are likely controlled by these factors. Plasmodesmata are important for numerous epidermal functions including cell-to-cell movement of TFs such as CPC and TRY (Schellmann et al., 2002; Wada et al., 2002) and transport of other macromolecules and metabolites (Lucas and Lee, 2004).

We also analyzed overrepresented ontology terms in the MYB77 and MYB44 targets that were classified as H cell-enriched genes. Among the MYB77 target genes in this category were known regulators of H cell fate, while numerous H cell-enriched MYB44 target genes were annotated as being involved in response to water and phosphate starvation (Figure 3.6C). The ontology category that was overrepresented in both target lists was *negative regulation of transcription* (6 MYB77 targets and 7 MYB44 targets), suggesting that these factors exert additional specific effects on the H cell transcriptome by regulating a subset of potentially repressive TFs.

The fact that MYB77 and MYB44 target a large number of genes that show H cell-enriched expression suggests that these factors serve as activators of transcription at these targets, and this is supported by published accounts of transcriptional control by these factors (Persak and Pitzschke, 2014).

However, both factors also target NH cell-enriched genes as well as genes without preferential expression between the cell types. This phenomenon was also observed for the H-enriched TFs ABI5, MYB33, and NAC083 (Figure 3.5), suggesting that certain TFs may generally serve as activators but may also have context-dependent repressive functions. Such a functional switch could occur through direct mechanisms such as structural alteration by alternative splicing or post-translational modification, functional alteration by partnering with a specific TF or chromatin-modifying complex, or perhaps indirectly by binding to a target site to occlude the binding of other factors necessary for transcriptional activation. The numerous reports of dual function transcription factors in animals and plants support the notion that this may be a general phenomenon (Ikeda et al., 2009; Boyle and Despres, 2010; Li et al., 2016b).

Collectively these results suggest that the MYB44/MYB77 module in the H cell specifies a cascade of downstream transcriptional regulation, some of which is positive and some of which is negative. This module likely represents an important hub in controlling H cell fate as well as a variety of physiological functions and environmental responses in this cell type. The fact that MYB77 was also discovered in our analyses of root tip TFs suggests that this factor likely has a broader role in other cell types during early root development, in addition to a role in specification of the H cell versus the NH cell fate. An important next step will be to perform genetic manipulations of these factors (knockout and inducible overexpression, for example), in order to test and elaborate on the specific predictions made by our model.

SUMMARY AND CONCLUSIONS

In this study, we used ATAC-seq profiling of accessible chromatin to investigate questions regarding the transcriptional regulatory landscape of plant genomes and its conservation across species. We also investigated the similarities and differences in open chromatin landscapes in two root cell types that arise from a common progenitor, allowing us to identify and analyze TFs that act specifically in one cell type versus the other. Overall, we are able to gain several new insights from this work.

In optimization of our ATAC-seq procedures, we found that the assay can be performed effectively on crudely purified nuclei but that this approach is limited by the large proportion of reads arising from

organelle genomes (Table 3.1). This issue is ameliorated by the use of the INTACT system to affinity-purify nuclei for ATAC-seq, which also provides access to individual cell types. Consistent with previous reports, we found that the data derived from ATAC-seq are highly similar to those from DNase-seq (Figure 3.1). In comparing our root tip ATAC-seq data to DNase-seq data from whole roots, we found that some hypersensitive regions were detected in one assay but not the other. This discrepancy is most likely attributable to differences in starting tissue and laboratory conditions, rather than biological differences in the chromatin regions sensitive to DNase I versus the hyperactive Tn5 transposase. This interpretation would fit with the large number of differences also observed in THS overlap between *Arabidopsis* root tip and epidermal cell types.

In a comparison of open chromatin among the root tip epigenomes of *Arabidopsis*, *Medicago*, tomato, and rice, we found the genomic distribution of THSs in each were highly similar. About 75% of THSs lie outside of transcribed regions, and the majority of these THSs are found within 3 kb upstream of the TSS in all species (Figure 3.2). Thus, the distance of upstream THSs from the TSS is relatively consistent among species and is not directly proportional to genome size or intergenic space for these representative plant species. Among genes with an upstream THS, 70% of these genes in *Arabidopsis*, *Medicago*, and rice have a single such feature, 20% have two upstream THSs, and less than 10% have three or more. In contrast, only 27% of tomato genes with an upstream THS have a single THS, 20% have two, and the proportion with 4-10 THSs is 2-7 times higher than that for any other species examined. This increase in THS number in tomato could be reflective of an increase in the number of regulatory elements per gene, but is perhaps more likely a result of the greater number of long-terminal repeat retrotransposons near genes in this species (Xu and Du, 2014). In either case, our investigation revealed that open chromatin sites – and by extension transcriptional regulatory elements – in all four species are focused in the TSS-proximal upstream regions and are relatively few in number per gene. This suggests that transcriptional regulatory elements in plants are generally fewer in number and are closer to the genes they regulate than those of animal genomes. For example, the median distance from an enhancer to its target TSSs in *Drosophila* was found to be 10 kb, and it was estimated that each gene had an average of four enhancers

(Kvon, Kazmar et al. 2014). It was also recently reported that in human T cells, the median distance between enhancers and promoters was 130 kb, far greater than the distances we have observed here across plant species (Mumbach, Satpathy et al. 2017).

Analysis of over-represented TF motifs in THSs across species suggested that many of the same TFs are at play in early root development in all species. Perhaps more surprisingly, co-regulation of specific gene sets by multiple TFs seems to be frequently maintained across species (Figure 3.3). Taken together with the lack of shared open chromatin profiles among orthologous genes and expressologs, these findings suggest that transcriptional regulatory elements may relocate over evolutionary time within a window of several kilobases upstream of the TSS, but regulatory control by specific TFs is relatively stable.

Our comparison of the two *Arabidopsis* root epidermal cell types, the hair (H) and non-hair (NH) cells, revealed that open chromatin profiles were highly similar between cell types. By examining THSs that were exclusive to one cell type, we were able to find several thousand THSs that were quantitatively more accessible in each cell type compared to the other (Figure 3.4). Mapping of these differential THSs (dTHSs) to their nearest genes revealed that in each cell type there were many dTHSs that were near genes expressed more abundantly in that cell type, as well as many near genes with the opposite expression pattern. This suggests that some dTHSs represented transcriptional activating events whereas others were repressive in nature.

Analysis of TF motifs at these dTHSs between cell types identified a suite of TFs that were more highly expressed in H cells and whose motifs were significantly overrepresented in H cell-enriched dTHSs. Analysis of three of these TFs – ABI5, MYB33, and NAC083 – revealed that each factor targets a large number of H cell-enriched genes as well as a smaller number of NH cell-enriched genes (Figure 3.5). These factors also have many overlapping target genes among them, and ABI5 and MYB33 both target seven additional H cell-enriched TFs. Among these seven H-enriched TFs are two additional MYB factors: MYB77 and MYB44 (Figure 3.6). Examination of the high confidence target genes of MYB77 and MYB44 revealed that these paralogous factors appeared to regulate each other as well as many other common target genes, including large numbers of other TF genes. Hundreds of the MYB77 and MYB44 target genes were

also more highly expressed in the H cell relative to the NH cell, suggesting that these factors set off a broad transcriptional cascade in the H cell type. In addition, they appear to directly regulate many H cell-enriched genes involved in cell fate specification and water and phosphate acquisition. This type of cooperative action by pairs of MYB paralogs has also been documented recently in *Arabidopsis* and other species (Millar and Gubler, 2005; Matus et al., 2017; Wang et al., 2017), and the fact that many target genes for each MYB factor are not regulated by the other may reflect a degree of subfunctionalization between the paralogs.

An important question arising from our results is whether classifying a TF as strictly an activator or repressor is generally accurate in most cases. For example, the H cell-enriched TFs that we examined all have apparent target genes that are highly expressed in the H cell type as well as targets that are expressed at very low levels, if at all, in the H cell type. In fact, these latter genes are often much more highly expressed in the NH cell type. Given that a number of these TFs have been shown to activate transcription in specific cases, this suggests that they promote the transcription of H cell-enriched targets and either repress or have no effect on NH cell-enriched target genes. One explanation for this phenomenon is that these TFs have "dual functionality" as activators and repressors, depending on the context (Bauer et al., 2010). However, it is equally possible that these factors do not play a direct role in gene repression. For example, the binding of an activator near a repressed gene may be functionally irrelevant to the regulation of that gene, or it may be the case that other gene-specific repressors may also be bound nearby and override the activity of the activator. This phenomenon will be worth exploring as it may deepen our understanding of the intricacies of transcriptional control.

In this study, we outline a widely applicable approach for combining chromatin accessibility profiling with available genome-wide binding data to construct models of TF regulatory networks. The putative TF regulatory pathways we have illuminated through our comparison across species and cell types provide important hypotheses regarding the evolution of gene regulatory mechanisms in plants and the mechanisms of cell fate specification, that are now open to experimental analysis.

METHODS

Plant materials and growth conditions

Plants used in this study were of the *Arabidopsis thaliana* Col-0 ecotype, the A17 ecotype of *Medicago truncatula*, the M82 LA3475 cultivar of tomato (*Solanum lycopersicum*), and the Nipponbare cultivar of rice (*Oryza sativa*). Transgenic plants of each species for INTACT were produced by transformation with a binary vector carrying both a constitutively expressed biotin ligase and constitutively expressed nuclear tagging fusion protein (NTF) containing a nuclear outer membrane association domain (Ron et al., 2014). The binary vector used for *Medicago* was identical to the tomato vector (Ron et al., 2014), but was constructed in a pB7WG vector containing phosphinothricin resistance gene for plant selection and it retains the original *AtACT2p* promoter. The binary vector used for rice is described in Reynoso *et al.* (submitted). Transformation of rice was carried out at UC Riverside and tomato transformation was carried out at the UC Davis plant transformation facility. *Arabidopsis* plants were transformed by the floral dip method (Clough and Bent, 1998) and composite transgenic *Medicago* plants were produced according to established procedures (Limpens et al., 2004).

For root tip chromatin studies, constitutive INTACT transgenic plant seeds were surface sterilized and sown on ½-strength Murashige and Skoog (MS) media (Murashige and Skoog, 1962) with 1% (w/v) sucrose in 150 mm diameter Petri plates, except for tomato and rice where full-strength MS with 1% (w/v) sucrose and without vitamins was used. Seedlings were grown on vertically oriented plates in controlled growth chambers for 7 days after germination, at which point the 1 cm root tips were harvested and frozen immediately in liquid N_2 for subsequent nuclei isolation. The growth temperature and light intensity was 20°C and 200 μmol/m_2/sec for *Arabidopsis* and *Medicago*, 23°C and 80 μmol/m_2/sec for tomato, and 28°C/25°C day/night and 110 μmol/m_2/sec for rice. Light cycles were 16 h light/8 h dark for all species.

For studies of the *Arabidopsis* root hair and non-hair cell types, previously described INTACT transgenic lines were used (Deal and Henikoff, 2010). These lines are in the Col-0 background and carry a constitutively expressed biotin ligase gene (*ACT2p:BirA*) and a transgene conferring cell type-specific

expression of the NTF gene (from the *GLABRA2* promoter in non-hair cells or the *ACTIN DEPOLYMERIZING FACTOR8* promoter in root hair cells). Plants were grown vertically on plates as described above for 7 days, at which point 1.25 cm segments from within the fully differentiated cell zone were harvested and flash frozen in liquid N_2. This segment of the root contains only fully differentiated cells and excludes the root tip below and any lateral roots above.

Nuclei isolation

For comparison of ATAC-seq using crude and INTACT-purified *Arabidopsis* nuclei, a constitutive INTACT line was used (*ACT2p:BirA*/*UBQ10p:NTF*) (Sullivan et al., 2014) and nuclei were isolated as described previously (Bajic et al., 2017). In short, after growth and harvesting as described above, 1-3 g of root tips were ground to a powder in liquid N_2 in a mortar and pestle and then resuspended in 10 ml of NPB (20 mM MOPS [pH 7], 40 mM NaCl, 90 mM KCl, 2 mM EDTA, 0.5 mM EGTA, 0.5 mM spermidine, 0.2 mM spermine, 1× Roche Complete protease inhibitors) with further grinding. This suspension was then filtered through a 70 μM cell strainer and centrifuged at 1,200 x g for 10 min at 4° C. After decanting, the nuclei pellet was resuspended in 1 ml of NPB and split into two 0.5 ml fractions in new tubes. Nuclei from one fraction were purified by INTACT using streptavidin-coated magnetic beads as previously described (Bajic et al., 2017) and kept on ice prior to counting and subsequent transposase integration reaction. Nuclei from the other fraction were purified by non-ionic detergent lysis of organelles and sucrose sedimentation, as previously described (Bajic et al., 2017). Briefly, these nuclei in 0.5 ml of NPB were pelleted at 1,200 x g for 10 min at 4° C, decanted, and resuspended thoroughly in 1 ml of cold EB2 (0.25 M sucrose, 10 mM Tris [pH 8], 10 mM $MgCl_2$, 1% Triton X-100, and 1× Roche Complete protease inhibitors). Nuclei were then pelleted at 1,200 x g for 10 min at 4° C, decanted, and resuspended in 300 μl of EB3 (1.7 M sucrose, 10 mM Tris [pH 8], 2 mM $MgCl_2$, 0.15% Triton X-100, and 1× Roche Complete protease inhibitors). This suspension was then layered gently on top of 300 μl of fresh EB3 in a 1.5 ml tube and centrifuged at 16,000 x g for 10 minutes at 4° C. Pelleted nuclei were then resuspended in 1 ml of cold NPB and kept on ice prior to counting and transposase integration.

For INTACT purification of total nuclei from root tips of *Medicago*, tomato and rice, as well as purification of *Arabidopsis* root hair and non-hair cell nuclei, 1-3 g of starting tissue was used. In all cases, nuclei were purified by INTACT and nuclei yields were quantified as described previously (Bajic et al., 2017).

Assay for transposase-accessible chromatin with sequencing (ATAC-seq)

Freshly purified nuclei to be used for ATAC-seq were kept on ice prior to the transposase integration reaction and never frozen. Transposase integration reactions and sequencing library preparations were then carried out as previously described (Bajic et al., 2017). In brief, 50,000 purified nuclei or 50 ng of *Arabidopsis* leaf genomic DNA were used in each 50 µl transposase integration reaction for 30 min at 37° C using Nextera reagents (Illumina, FC-121-1030). DNA fragments were purified using the Minelute PCR purification kit (Qiagen), eluted in 11 µl of elution buffer, and the entirety of each sample was then amplified using High Fidelity PCR Mix (NEB) and custom barcoded primers for 9-12 total PCR cycles. These amplified ATAC-seq libraries were purified using AMPure XP beads (Beckman Coulter), quantified by qPCR with the NEBNext Library Quantification Kit (NEB), and analyzed on a Bioanalyzer High Sensitivity DNA Chip (Agilent) prior to pooling and sequencing.

High throughput sequencing

Sequencing was carried out using the Illumina NextSeq 500 or HiSeq2000 instrument at the Georgia Genomics Facility at the University of Georgia. Sequencing reads were either single-end 50 nt or paired-end 36 nt and all libraries that were to be directly compared were pooled and sequenced on the same flow cell.

Sequence read mapping, processing, and visualization

Sequencing reads were mapped to their corresponding genome of origin using Bowtie2 software (Langmead and Salzberg, 2012) with default parameters. Genome builds used in this study were *Arabidopsis* version TAIR10, *Medicago* version Mt4.0, Tomato version SL2.4, and Rice version IRGSP 1.0.30. Mapped reads in *.sam* format were converted to *.bam* format and sorted using Samtools 0.1.19 (Li et al., 2009). Mapped reads were then filtered using Samtools to retain only those reads with a mapping quality score of 2 or higher (Samtools "*view*" command with option "*-q 2*" to set mapping quality cutoff). *Arabidopsis* ATAC-seq reads were further filtered with Samtools to remove those mapping to either the chloroplast or mitochondrial genomes, and root hair and non-hair cell datasets were also subsampled such that the experiments within a biological replicate had the same number of mapped reads prior to further analysis. For normalization and visualization, the filtered, sorted *.bam* files were converted to bigwig format using the "*bamcoverage*" script in deepTools 2.0 (Ramirez et al., 2016) with a bin size of 1 bp and RPKM normalization. Use of the term *normalization* in this paper refers to this process. Heatmaps and average plots displaying ATAC-seq data were also generated using the "*computeMatrix*" and "*plotHeatmap*" functions in the deepTools package. Genome browser images were made using the Integrative Genomics Viewer (IGV) 2.3.68 (Thorvaldsdottir et al., 2013) with bigwig files processed as described above.

Identification of orthologous genes among species

Orthologous genes among species were selected exclusively from syntenic regions of the four genomes. Syntenic orthologs were identified using a combination of CoGe SynFind (https://genomevolution.org/CoGe/SynFind.pl) with default parameters, and CoGe SynMap (https://genomevolution.org/coge/SynMap.pl) with QuotaAlign feature selected and a minimum of six aligned pairs required (Lyons and Freeling 2008, Lyons, Pedersen et al. 2008).

Peak calling to detect transposase hypersensitive sites (THSs)

Peak calling on ATAC-seq data was performed using the "*Findpeaks*" function of the HOMER package (Heinz et al., 2010). The parameters "*-region*" and "*-minDist 150*" were used to allow

identification of variable length peaks and to set a minimum distance of 150 bp between peaks before they are merged into a single peak, respectively. We refer to the peaks called in this way as “transposase hypersensitive sites”, or THSs.

Genomic distribution of THSs

For each genome, the distribution of THSs relative to genomic features was assessed using the PAVIS web tool (Huang et al., 2013) with “upstream” regions set as the 2,000 bp upstream of the annotated transcription start site and “downstream” regions set as 1,000 bp downstream of the transcript termination site.

Transcription factor motif analyses

ATAC-seq transposase hypersensitive sites (THSs) that were found in two replicates of each sample were used for motif analysis. The regions were adjusted to the same size (500 bp for root tip THSs or 300 bp for cell type-specific dTHSs). The MEME-ChIP pipeline (Machanick and Bailey, 2011) was run on the repeat-masked fasta files representing each THS set to identify overrepresented motifs, using default parameters. For further analysis, we used the motifs derived from the DREME, MEME, and CentriMo programs that were significant matches (E value < 0.05) to known motifs. Known motifs from both Cis-BP (Weirauch, Yang et al. 2014) and the DAP-seq database (O'Malley, Huang et al. 2016) were used in all motif searches.

Assignment of THSs to genes

For each ATAC-seq data set the THSs were assigned to genes using the “*TSS*” function of the PeakAnnotator 1.4 program (Salmon-Divon et al., 2010). This program assigns each peak/THS to the closest transcription start site (TSS), whether upstream or downstream, and reports the distance from the peak center to the TSS based on the genome annotations described above.

ATAC-seq footprinting

To examine motif-centered footprints for TFs of interest we used the "*dnase_average_profile.py"* script in the pyDNase package (Piper et al., 2013). The script was used in ATAC-seq mode ["*-A*" parameter] with otherwise default parameters.

Publicly available DNase-seq, DAP-seq, ChIP-seq, and RNA-seq data

For comparison to our ATAC-seq data from root tips, we used a published DNase-seq dataset from 7-day-old whole *Arabidopsis* roots (SRX391990), which was generated from the same INTACT transgenic line used in our experiments (Sullivan et al., 2014).

Publicly available ChIP-seq and DAP-seq datasets were also used to identify genomic binding sites for transcription factors of interest. These include ABF3 (AT4G34000; SRX1720080) and MYB44 (AT5G67300; SRX1720040)(Song et al., 2016), HY5 (AT5G11260; SRX1412757), CBF2 (AT4G25470; SRX1412036), MYB77 (AT3G50060; SRX1412453), ABI5 (AT2G36270; SRX670505), MYB33 (AT5G06100; SRX1412418), NAC083 (AT5G13180; SRX1412546), MYB77 (AT3G50060; SRX1412453), WRKY27 (AT5G52830; SRX1412681), and At5g04390 (SRX1412214) (O'Malley et al., 2016). Raw reads from these files were mapped and processed as described above for ATAC-seq data, including peak calling with the HOMER package.

Published RNA-seq data from *Arabidopsis* root hair and non-hair cells (Li et al., 2016a) were used to define transcripts that were specifically enriched in the root hair cell relative to the non-hair cell (hair cell enriched genes), and vice versa (non-hair enriched genes). We defined cell type-enriched genes as those whose transcripts were at least two-fold more abundant in one cell type than the other and had an abundance of at least five RPKM in the cell type with higher expression.

Defining high confidence target sites for transcription factors

We used FIMO (Grant, Bailey et al. 2011) to identify motif occurrences for TFs of interest, and significant motif occurrences were considered to be those with a p-value < 0.0001. Genome-wide high

confidence binding sites for a given transcription factor were defined as transposase hypersensitive sites in a given cell type or tissue that also contain a significant motif occurrence for the factor and also overlap with a known enriched region for that factor from DAP-seq or ChIP-seq data (see also Figure S3.2 for a schematic diagram of this process).

Gene ontology analysis

Gene Ontology (GO) analyses using only *Arabidopsis* genes were carried out using the GeneCodis 3.0 program (Nogales-Cadenas et al., 2009; Tabas-Madrid et al., 2012). Hypergeometric tests were used with p-value correction using the false discovery rate (FDR) method. AgriGO was used for comparative GO analysis of gene lists among species, using default parameters (Du et al., 2010; Tian et al., 2017).

Accession Numbers

The raw and processed ATAC-seq data described here have been deposited to the NCBI Gene Expression Omnibus (GEO) database under record number GSE101482. The characteristics of each dataset (individual accession number, read numbers and mapping characteristics, and THS statistics) are included in Table S3.8.

ACKNOWLEDGEMENTS

We thank Paja Sijacic and Shannon Torres for constructive criticism of the manuscript. This work was supported by funding from the National Science Foundation (Plant Genome Research Program grant #IOS-123843) to J.B-S., N.S., S.M.B., and R.B.D.; D.A.W. was supported in part by funding from the Elise Taylor Stocking Memorial Fellowship, and K.K. was supported in part by the Finnish Cultural Foundation.

AUTHOR CONTRIBUTIONS

R.B.D., S.M.B., N.S., J.B-S., K.A.M., M. B., K.K, and M.R, G.P., and D.A.W. designed the research project. K.A.M. performed all experiments on *Arabidopsis* root tips as well as hair and non-hair

cells. M.B. performed all experiments on *Medicago* root tips, K.K., D.A.W, and K.Z. performed all experiments on tomato root tips, and M.R. and G.P. performed all experiments on rice root tips. M.W. performed all analyses of syntenic regions and identification of orthologous genes among species. K.B., M.D., and C.Q. analyzed ATAC-seq data sets with Hotspot software and also contributed expertise in other analyses. R.B.D, K.A.M, and M.B. analyzed the data, and R.B.D. drafted the manuscript with subsequent input and editing from all authors.

FIGURES

Figure 3.1

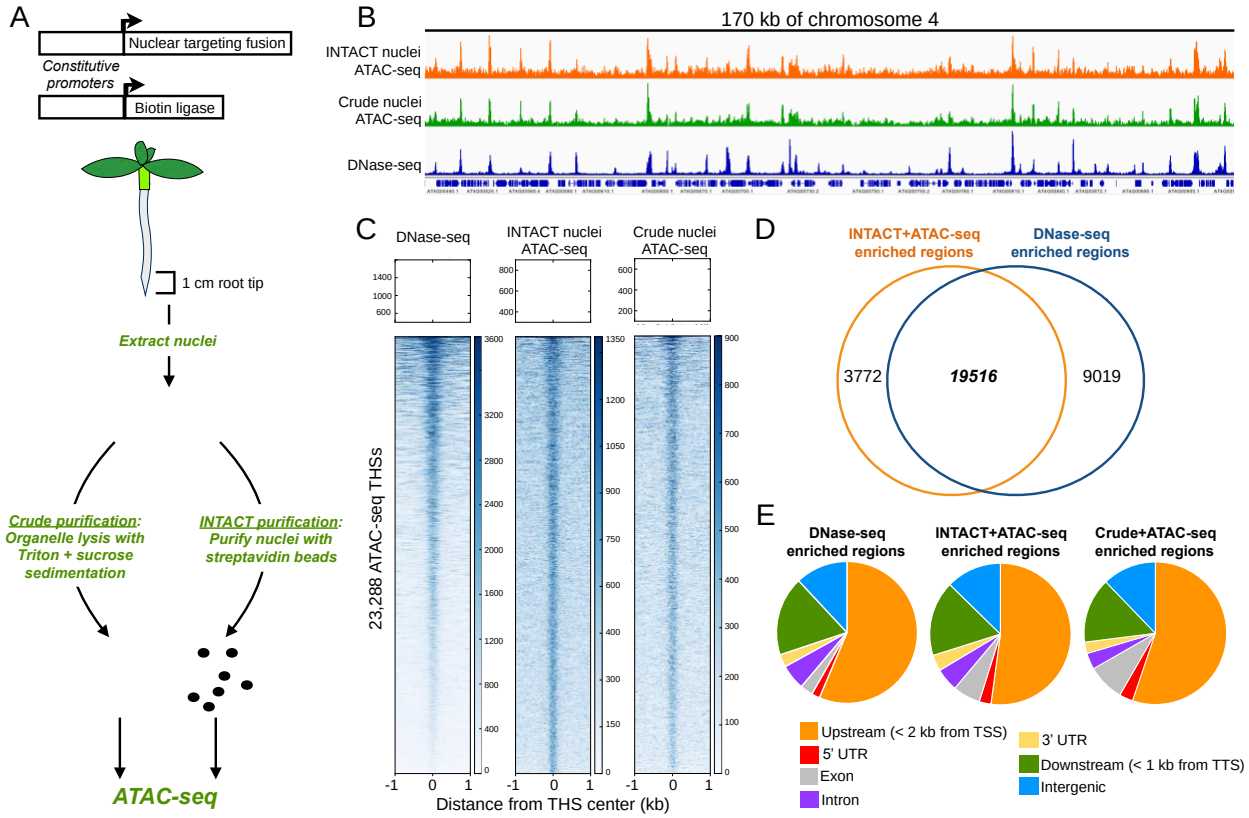


Figure 3.1 Application of ATAC-seq to *Arabidopsis* and comparison with DNase-seq data. (A) Schematic of the INTACT system and strategy for testing ATAC-seq on nuclei with different levels of purity. Upper panel shows the two transgenes used in the INTACT system: the nuclear targeting fusion (NTF) and biotin ligase. Driving expression of both transgenes using constitutive promoters generates biotinylated nuclei in all cell types. Below is a diagram of a constitutive INTACT transgenic plant, showing the 1 cm root tip section used for all nuclei purifications. Root tip nuclei were isolated from transgenic plants and either purified by detergent lysis of organelles followed by sucrose sedimentation (Crude) or purified using streptavidin beads (INTACT). In each case 50,000 purified nuclei were used as input for ATAC-seq. **(B)** Genome browser shot of ATAC-seq data along a 170 kb stretch of chromosome 4 from INTACT-purified and Crude nuclei, as well as DNase-seq data from whole root tissue. Gene models are displayed on the bottom track. **(C)** Average plots and heatmaps of DNase-seq and ATAC-seq signals at the

23,288 ATAC-seq transposase hypersensitive sites (THSs) in the INTACT-ATAC-seq dataset. The regions in the heatmaps are ranked from highest DNase-seq signal (top) to lowest (bottom) **(D)** Venn diagram showing the overlap of enriched regions identified in root tip INTACT-ATAC-seq and whole root DNase-seq datasets. **(E)** Genomic distributions of enriched regions identified in DNase-seq, INTACT-ATAC-seq, and Crude-ATAC-seq datasets.

Figure 3.2

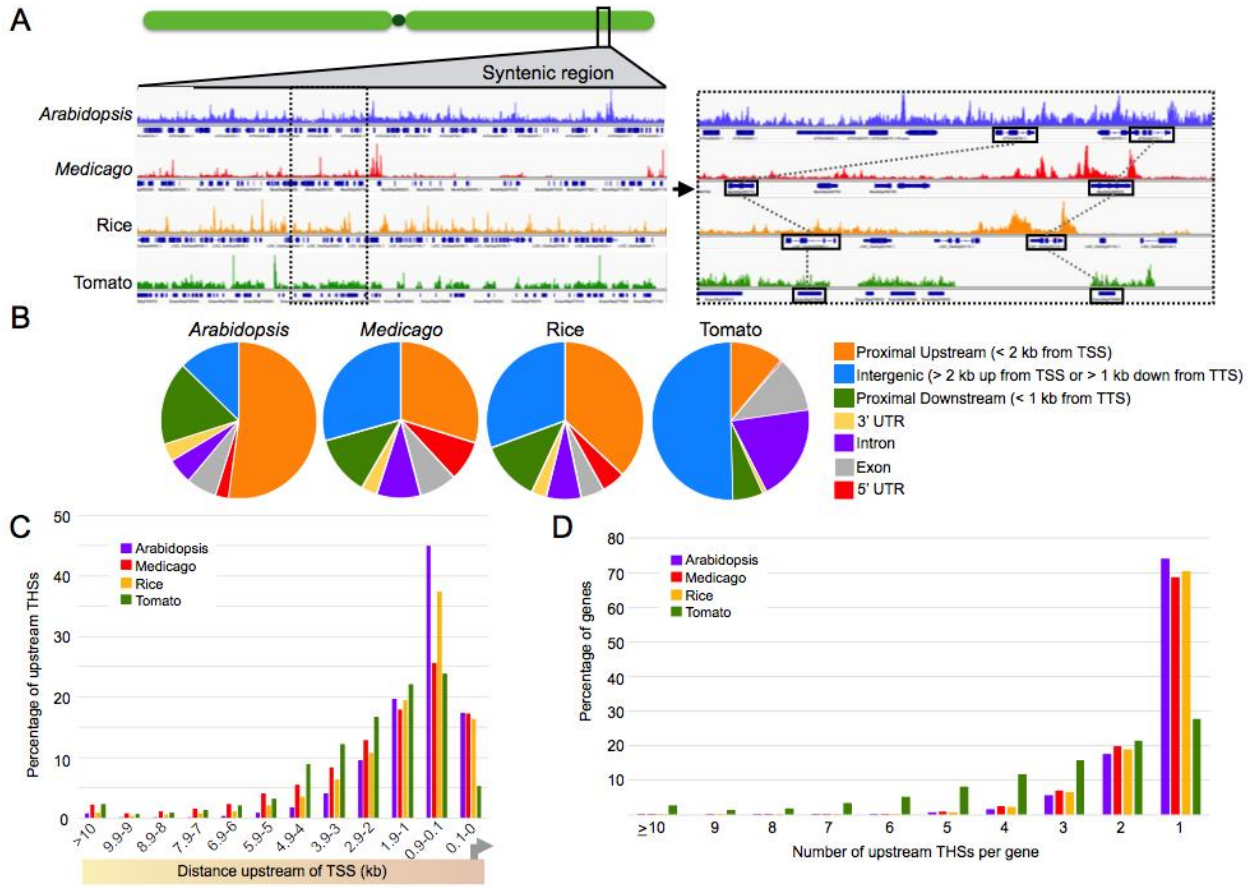


Figure 3.2 ATAC-seq profiling of *Arabidopsis*, *Medicago*, tomato, and rice. (A) Comparison of ATAC-seq data along syntenic regions across the species. The left panel shows a genome browser shot of ATAC-seq data across a syntenic region of all four genomes. ATAC-seq data tracks are shown above the corresponding gene track for each species. The right panel is an enlargement of the region surrounded by a dotted box in the left panel. Orthologous genes are surrounded by black boxes connected by dotted lines between species. Note the apparent similarity in transposase hypersensitivity upstream and downstream of the rightmost orthologs. **(B)** Distribution of ATAC-seq transposase hypersensitive sites (THSs) relative to genomic features in each species. **(C)** Distribution of upstream THSs relative to genes in each species. THSs are binned by distance upstream of the transcription start site (TSS). The number of peaks in each bin is expressed as a percentage of the total upstream THS number in that species. **(D)** Number of upstream

THSs per gene in each species. Graph shows the percentage of all genes with a given number of upstream THSs.

Figure 3.3

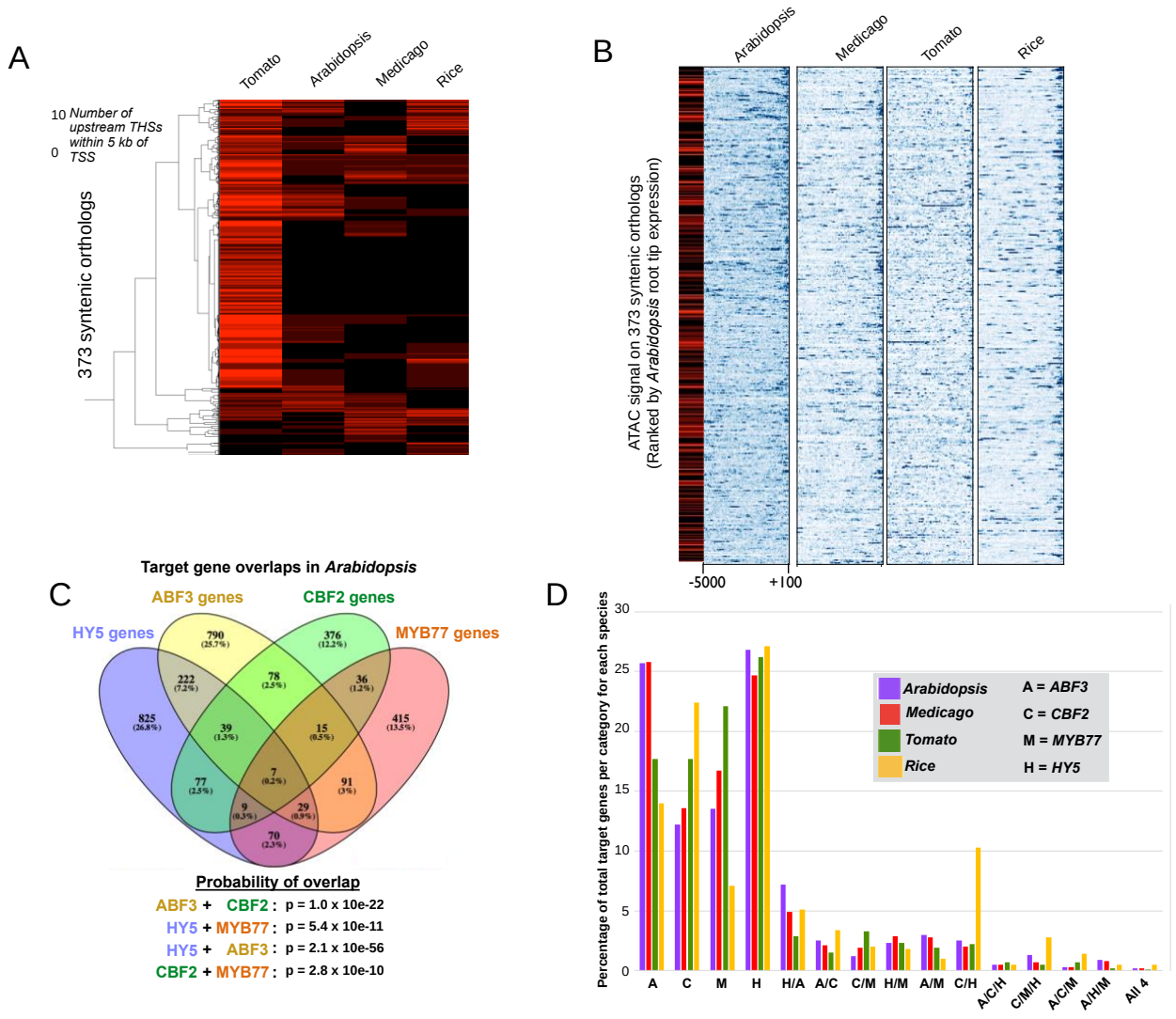


Figure 3.3 Characterization of open chromatin regions and regulatory elements in *Arabidopsis*, *Medicago*, tomato, and rice. (A) Heatmap showing the number of upstream THSs at each of 373 syntenic orthologs in each species. Each row of the heatmap represents a syntenic ortholog, and the number of THSs within 5 kb upstream of the TSS is indicated with a black-to-red color scale for each ortholog in each species. Hierarchical clustering was performed on orthologs using uncentered correlation and average linkage. **(B)** Normalized ATAC-seq signals upstream of orthologous genes. Each row of the heatmaps represents the upstream region of one of the 373 syntenic orthologs in each species. ATAC-seq signal is shown across each ortholog from +100 to -5000 bp relative to the TSS, where blue is high signal and white

is no signal. Heatmaps are ordered by transcript level of each *Arabidopsis* ortholog in the root tip, from highest (top) to lowest (bottom). The leftmost heatmap in black-to-red scale indicates the number of upstream THSs from -100 to -5000 bp associated with each of the *Arabidopsis* orthologs, on the same scale as in (A). **(C)** Overlap of predicted target genes for HY5, ABF3, CBF2, and MYB77 in the *Arabidopsis* root tip. Predicted binding sites for each factor are those THSs that also contain a significant motif occurrence for that factor. Venn diagram shows the numbers of genes with predicted binding sites for each factor alone and in combination with other factors. Significance of target gene set overlap between each TF pair was calculated using a hypergeometric test with a population including all *Arabidopsis* genes reproducibly associated with an ATAC-seq peak in the root tip (13,714 total genes). For each overlap, we considered all genes co-targeted by the two factors. **(D**) Conveying data similar to that in (C), the clustered bar graph shows the percentage of total target genes that fall into a given regulatory category (targeted by a single TF or combination of TFs) in each species.

Figure 3.4

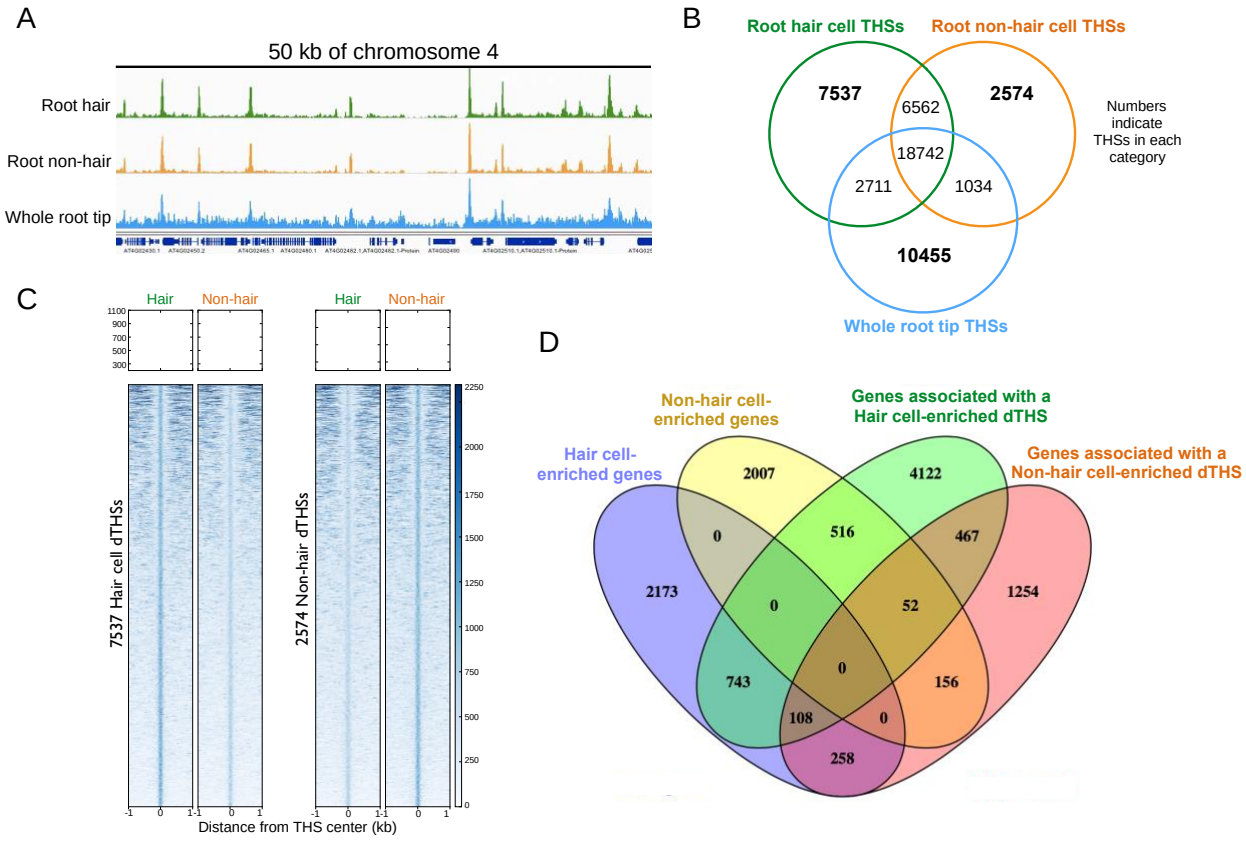


Figure 3.4 Characterization of open chromatin regions in the *Arabidopsis* root hair and non-hair cell types. (A) Genome browser shot of ATAC-seq data from root hair cell, non-hair cell, and whole root tip representing 50 kb of Chromosome 4. **(B)** Overlap of THSs found in two biological replicates of each cell type or tissue. Numbers in bold indicate THSs that are only found in a given cell type or tissue (differential THSs, or dTHSs). **(C)** Average plots and heatmaps showing normalized ATAC-seq signals over 7,537 root hair cell dTHSs (left panels) and 2,574 non-hair cell-enriched dTHSs (right panels). Heatmaps are ranked in decreasing order of total ATAC-seq signal in the hair cell panel in each comparison. Data from one biological replicate is shown here and both replicate experiments showed very similar results. **(D)** Venn diagram of overlaps between cell type-enriched gene sets and genes associated with cell type-enriched dTHSs. Transcriptome data from hair (purple) and non-hair cells (yellow) are from Li et al. (2016) *Developmental Cell*. Genes were considered cell type-enriched if they had a 2-fold or higher difference between cell types and a read count of 5 RPKM or greater in the cell type with higher expression.

Figure 3.5

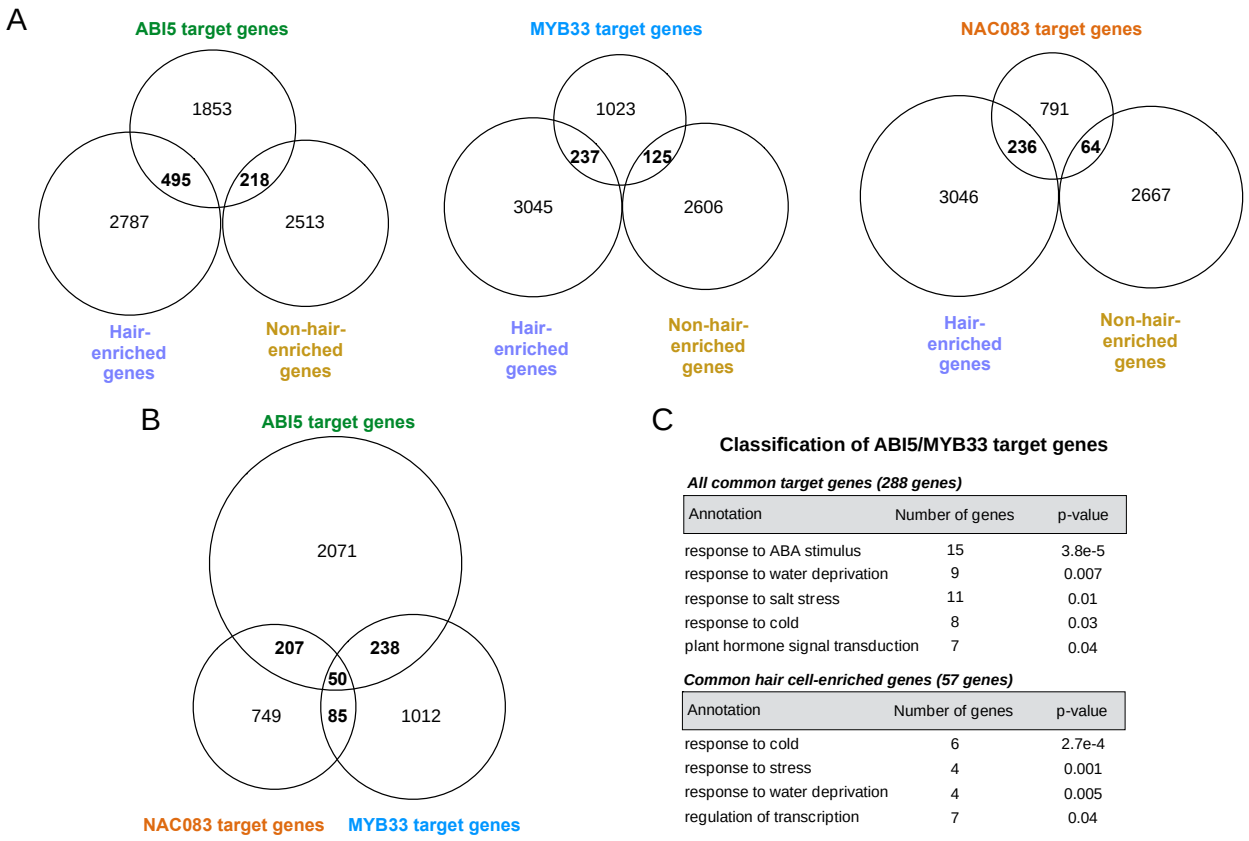
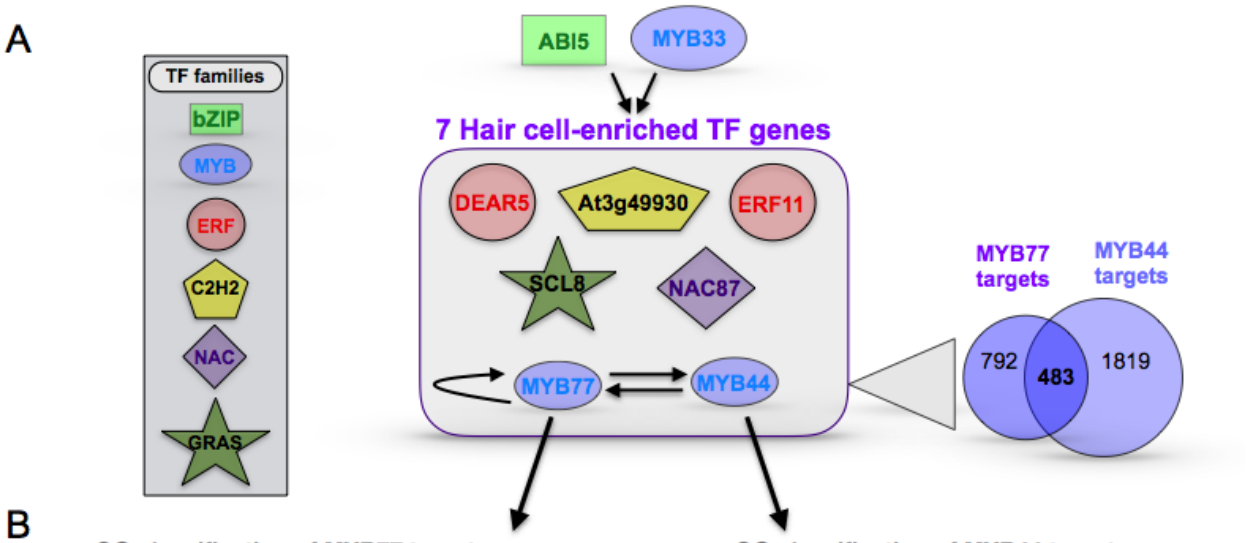


Figure 3.5 Targeting of cell type-enriched genes by H cell-enriched TFs, and co-regulatory associations among H-cell enriched TFs. Genome-wide high confidence binding sites for each TF were defined as open chromatin regions in the hair cell that contain a significant motif occurrence for the factor and also overlap with a known enriched region for that factor from DAP-seq or ChIP-seq data. Target genes were defined by assigning each high confidence binding site to the nearest TSS. **(A)** Venn diagrams showing high confidence target genes for ABI5, MYB33, and NAC083 and their overlap with cell type-enriched genes. **(B)** Overlap of ABI5, MYB33, and NAC083 high confidence target genes. **(C)** Gene Ontology (GO) analysis was performed to illuminate biological functions of genes co-targeted by ABI5 and MYB33. The upper panel shows significantly enriched GO terms for all 288 genes targeted by both ABI5 and MYB33. For each enriched annotation term, the number of genes in the set with that term is shown, followed by the FDR-corrected p-value. The lower panel lists significantly enriched GO-terms for the 57 hair cell-enriched genes co-targeted by ABI5 and MYB33. The seven hair cell-enriched genes

associated with the term *regulation of transcription* were chosen for further analysis. All annotation terms in the lists are at the Biological Process level except for the KEGG pathway term ‘plant hormone signal transduction’.

Figure 3.6



B

GO classification of MYB77 target genes

All target genes (1277 genes)

Annotation	Number of genes	p-value
Plasmodesma (CC)	66	4.5e-5
Response to salt stress (BP)	36	2.6e-4
Response to ethylene stimulus (BP)	17	2.7e-4
Transcription factor activity (MF)	102	6.1e-4
Response to SA stimulus (BP)	16	4.7e-3
Response to ABA stimulus (BP)	26	0.04

GO classification of MYB44 target genes

All target genes (2302 genes)

Annotation	Number of genes	p-value
Plasmodesma (CC)	99	5.1e-5
Response to salt stress (BP)	51	6.6e-4
Response to ethylene stimulus (BP)	23	2.6e-4
Transcription factor activity (MF)	183	7.9e-7
Response to SA stimulus (BP)	23	1.5e-3
Response to ABA stimulus (BP)	44	1.2e-3
Response to JA stimulus (BP)	25	1.0e-3
Response to Auxin stimulus (BP)	35	0.014
Biosynthesis of secondary metabolites (KEGG)	74	5.3e-3

C

Hair cell-enriched target genes (219 genes)

Annotation	Number of genes	p-value
Negative regulation of transcription (BP)	6	5.3e-3
Root hair cell development (BP)	3	3.7e-3
Root hair cell differentiation (BP)	2	0.043

Hair cell-enriched target genes (425 genes)

Annotation	Number of genes	p-value
Negative regulation of transcription (BP)	7	0.01
Response to water deprivation (BP)	14	7.8e-5
Response to phosphate starvation (BP)	6	0.02

Figure 3.6 A transcriptional regulatory module in the root hair cell type. (A) Diagram of the proposed regulatory module under control of ABI5 and MYB33. As referenced in **Figure 3.5C**, ABI5 and MYB33 co-target seven TFs that are preferentially expressed in the hair cell relative to the non-hair cell type. The family classification of each of the seven TFs is denoted in the figure key. Among the seven hair cell-specific target TFs are two MYB family members, MYB77 and MYB44. High confidence binding sites for these two MYB factors were again defined as open chromatin regions in the hair cell that contain a significant motif occurrence for the factor and also overlap with a known enriched region for that factor

from DAP-seq or ChIP-seq data. Each high confidence binding site was then assigned to the nearest TSS to define the target gene for that site. This analysis revealed that MYB44 and MYB77 target each other, and MYB77 targets itself. Both factors target thousands of additional genes, 483 of which are in common (Venn diagram on the lower right of the schematic. Arrows coming down from MYB77 and MYB44 point to GO analyses of that factor's target genes. **(B)** The upper tables represent enriched annotation terms for all target genes of the factor, regardless of differential expression between H and NH cells, while the lower tables **(C)** represent enrichment of terms within target genes that are preferentially expressed in the hair cell relative to the non-hair cell. Annotation term levels are indicated as Cellular Component (CC), Biological Process (BP), Molecular Function (MF) or KEGG pathway (KEGG). For each annotation, the number of target genes associated with that term is shown to the right of the term, followed by the FDR-corrected p-value for the term enrichment in the rightmost column. Groups of terms boxed in gray are those that differ between MYB44 and MYB77. The structure of the module suggests that ABI5 and MYB33 drive a cascade of TFs including MYB77 and MYB44, which act to amplify this signal and also further regulate many additional TFs. Additional target genes of MYB77 and MYB44 include hair cell differentiation factors, hormone response genes, secondary metabolic genes, and genes encoding components of important cellular structures such as plasmodesmata.

LITERATURE CITED

Abdeen, A., Schnell, J., and Miki, B. (2010). Transcriptome analysis reveals absence of unintended effects in drought-tolerant transgenic plants overexpressing the transcription factor ABF3. BMC Genomics **11,** 69.

Bajic, M., Maher, K.A., and Deal, R.B. (2017). Identification of Open Chromatin Regions in Plant Genomes Using ATAC-Seq. Methods in Molecular Biology **1675,** 183-201.

Bauer, D.C., Buske, F.A., and Bailey, T.L. (2010). Dual-functioning transcription factors in the developmental gene network of Drosophila melanogaster. BMC Bioinformatics **11,** 366.

Bonn, S., Zinzen, R.P., Girardot, C., Gustafson, E.H., Perez-Gonzalez, A., Delhomme, N., Ghavi-Helm, Y., Wilczynski, B., Riddell, A., and Furlong, E.E. (2012). Tissue-specific analysis of chromatin state identifies temporal signatures of enhancer activity during embryonic development. Nat Genet **44,** 148-156.

Boyle, P., and Despres, C. (2010). Dual-function transcription factors and their entourage: unique and unifying themes governing two pathogenesis-related genes. Plant Signal Behav **5,** 629-634.

Buenrostro, J.D., Wu, B., Chang, H.Y., and Greenleaf, W.J. (2015). ATAC-seq: A Method for Assaying Chromatin Accessibility Genome-Wide. Curr Protoc Mol Biol **109,** 21 29 21-29.

Buenrostro, J.D., Giresi, P.G., Zaba, L.C., Chang, H.Y., and Greenleaf, W.J. (2013). Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. Nature Methods **10,** 1213-1218.

Clough, S.J., and Bent, A.F. (1998). Floral dip: a simplified method for Agrobacterium-mediated transformation of Arabidopsis thaliana. Plant J **16,** 735-743.

Deal, R.B., and Henikoff, S. (2010). A simple method for gene expression and chromatin profiling of individual cell types within a tissue. Dev Cell **18,** 1030-1040.

Du, Z., Zhou, X., Ling, Y., Zhang, Z., and Su, Z. (2010). agriGO: a GO analysis toolkit for the agricultural community. Nucleic Acids Res **38,** W64-70.

Dubos, C., Stracke, R., Grotewold, E., Weisshaar, B., Martin, C., and Lepiniec, L. (2010). MYB transcription factors in Arabidopsis. Trends in plant science **15,** 573-581.

Finkelstein, R.R., and Lynch, T.J. (2000). The Arabidopsis abscisic acid response gene ABI5 encodes a basic leucine zipper transcription factor. Plant Cell **12,** 599-609.

Grant, C.E., Bailey, T.L., and Noble, W.S. (2011). FIMO: scanning for occurrences of a given motif. Bioinformatics **27,** 1017-1018.

Gross, D.S., and Garrard, W.T. (1988). Nuclease hypersensitive sites in chromatin. Annu Rev Biochem **57,** 159-197.

Heintzman, N.D., Hon, G.C., Hawkins, R.D., Kheradpour, P., Stark, A., Harp, L.F., Ye, Z., Lee, L.K., Stuart, R.K., Ching, C.W., Ching, K.A., Antosiewicz-Bourget, J.E., Liu, H., Zhang, X., Green, R.D., Lobanenkov, V.V., Stewart, R., Thomson, J.A., Crawford, G.E., Kellis, M., and Ren, B. (2009). Histone modifications at human enhancers reflect global cell-type-specific gene expression. Nature **459,** 108-112.

Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y.C., Laslo, P., Cheng, J.X., Murre, C., Singh, H., and Glass, C.K. (2010). Simple Combinations of Lineage-Determining Transcription Factors Prime cis-Regulatory Elements Required for Macrophage and B Cell Identities. Molecular Cell **38,** 576-589.

Henikoff, S. (2008). Nucleosome destabilization in the epigenetic regulation of gene expression. Nat Rev Genet **9,** 15-26.

Hesselberth, J.R., Chen, X., Zhang, Z., Sabo, P.J., Sandstrom, R., Reynolds, A.P., Thurman, R.E., Neph, S., Kuehn, M.S., Noble, W.S., Fields, S., and Stamatoyannopoulos, J.A. (2009). Global mapping of protein-DNA interactions in vivo by digital genomic footprinting. Nature Methods **6,** 283-289.

Huang, W., Loganantharaj, R., Schroeder, B., Fargo, D., and Li, L. (2013). PAVIS: a tool for Peak Annotation and Visualization. Bioinformatics **29,** 3097-3099.

Ikeda, M., Mitsuda, N., and Ohme-Takagi, M. (2009). Arabidopsis WUSCHEL is a bifunctional transcription factor that acts as a repressor in stem cell regulation and as an activator in floral patterning. Plant Cell **21,** 3493-3505.

Jaradat, M.R., Feurtado, J.A., Huang, D., Lu, Y., and Cutler, A.J. (2013). Multiple roles of the transcription factor AtMYBR1/AtMYB44 in ABA signaling, stress responses, and leaf senescence. BMC Plant Biol **13,** 192.

John, S., Sabo, P.J., Thurman, R.E., Sung, M.H., Biddie, S.C., Johnson, T.A., Hager, G.L., and Stamatoyannopoulos, J.A. (2011). Chromatin accessibility pre-determines glucocorticoid receptor binding patterns. Nat Genet **43,** 264-268.

Jung, C., Shim, J.S., Seo, J.S., Lee, H.Y., Kim, C.H., Choi, Y.D., and Cheong, J.J. (2010). Non-specific phytohormonal induction of AtMYB44 and suppression of jasmonate-responsive gene activation in Arabidopsis thaliana. Mol Cells **29,** 71-76.

Kang, J.Y., Choi, H.I., Im, M.Y., and Kim, S.Y. (2002). Arabidopsis basic leucine zipper proteins that mediate stress-responsive abscisic acid signaling. Plant Cell **14,** 343-357.

Kerr, T.C., Abdel-Mageed, H., Aleman, L., Lee, J., Payton, P., Cryer, D., and Allen, R.D. (2017). Ectopic expression of two AREB/ABF orthologs increases drought tolerance in cotton (Gossypium hirsutum). Plant Cell Environ.

Knight, H., Zarka, D.G., Okamoto, H., Thomashow, M.F., and Knight, M.R. (2004). Abscisic acid induces CBF gene transcription and subsequent induction of cold-regulated genes via the CRT promoter element. Plant Physiol **135,** 1710-1717.

Kvon, E.Z., Kazmar, T., Stampfel, G., Yanez-Cuna, J.O., Pagani, M., Schernhuber, K., Dickson, B.J., and Stark, A. (2014). Genome-scale functional characterization of Drosophila developmental enhancers in vivo. Nature **512,** 91-95.

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

Lee, T.I., and Young, R.A. (2000). Transcription of eukaryotic protein-coding genes. Annu Rev Genet **34,** 77-137.

Li, D., Li, Y., Zhang, L., Wang, X., Zhao, Z., Tao, Z., Wang, J., Wang, J., Lin, M., Li, X., and Yang, Y. (2014). Arabidopsis ABA Receptor RCAR1/PYL9 Interacts with an R2R3-Type MYB Transcription Factor, AtMYB44. International Journal of Molecular Sciences **15,** 8473-8490.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., and Genome Project Data Processing, S. (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics **25,** 2078-2079.

Li, J., Farmer, A.D., Lindquist, I.E., Dukowic-Schulze, S., Mudge, J., Li, T., Retzel, E.F., and Chen, C. (2012). Characterization of a set of novel meiotically-active promoters in Arabidopsis. BMC Plant Biol **12,** 104.

Li, S., Yamada, M., Han, X., Ohler, U., and Benfey, P.N. (2016a). High-Resolution Expression Map of the Arabidopsis Root Reveals Alternative Splicing and lincRNA Regulation. Developmental Cell **39,** 508-522.

Li, T., Wu, X.Y., Li, H., Song, J.H., and Liu, J.Y. (2016b). A Dual-Function Transcription Factor, AtYY1, Is a Novel Negative Regulator of the Arabidopsis ABA Response Network. Mol Plant **9,** 650-661.

Limpens, E., Ramos, J., Franken, C., Raz, V., Compaan, B., Franssen, H., Bisseling, T., and Geurts, R. (2004). RNA interference in Agrobacterium rhizogenes-transformed roots of Arabidopsis and Medicago truncatula. J Exp Bot **55,** 983-992.

Lu, Z., Hofmeister, B.T., Vollmers, C., DuBois, R.M., and Schmitz, R.J. (2017). Combining ATAC-seq with nuclei sorting for discovery of cis-regulatory regions in plant genomes. Nucleic Acids Res **45,** e41.

Lucas, W.J., and Lee, J.Y. (2004). Plasmodesmata as a supracellular control network in plants. Nat Rev Mol Cell Biol **5,** 712-726.

Ludwig, M.Z., Bergman, C., Patel, N.H., and Kreitman, M. (2000). Evidence for stabilizing selection in a eukaryotic enhancer element. Nature **403,** 564-567.

Ludwig, M.Z., Palsson, A., Alekseeva, E., Bergman, C.M., Nathan, J., and Kreitman, M. (2005). Functional evolution of a cis-regulatory module. PLoS Biol **3,** e93.

Lyons, E., and Freeling, M. (2008). How to usefully compare homologous plant genes and chromosomes as DNA sequences. Plant J **53,** 661-673.

Lyons, E., Pedersen, B., Kane, J., and Freeling, M. (2008). The Value of Nonmodel Genomes and an Example Using SynMap Within CoGe to Dissect the Hexaploidy that Predates the Rosids. Tropical Plant Biology **1,** 181-190.

Ma, W., Noble, W.S., and Bailey, T.L. (2014). Motif-based analysis of large nucleotide data sets using MEME-ChIP. Nature protocols **9,** 1428-1450.

Machanick, P., and Bailey, T.L. (2011). MEME-ChIP: motif analysis of large DNA datasets. Bioinformatics **27,** 1696-1697.

Masucci, J.D., Rerie, W.G., Foreman, D.R., Zhang, M., Galway, M.E., Marks, M.D., and Schiefelbein, J.W. (1996). The homeobox gene GLABRA2 is required for position-dependent cell differentiation in the root epidermis of Arabidopsis thaliana. Development **122,** 1253-1260.

Mathelier, A., Zhao, X., Zhang, A.W., Parcy, F., Worsley-Hunt, R., Arenillas, D.J., Buchman, S., Chen, C.Y., Chou, A., Ienasescu, H., Lim, J., Shyr, C., Tan, G., Zhou, M., Lenhard, B., Sandelin, A., and Wasserman, W.W. (2014). JASPAR 2014: an extensively expanded and updated open-access database of transcription factor binding profiles. Nucleic Acids Res **42,** D142-147.

Matus, J.T., Cavallini, E., Loyola, R., Holl, J., Finezzo, L., Dal Santo, S., Vialet, S., Commisso, M., Roman, F., Schubert, A., Alcalde, J.A., Bogs, J., Ageorges, A., Tornielli, G.B., and Arce-Johnson, P. (2017). A Group of Grapevine MYBA Transcription Factors Located in Chromosome 14 Control Anthocyanin Synthesis in Vegetative Organs with Different Specificities Compared to the Berry Color Locus. Plant J.

Medford, J.I., Elmer, J.S., and Klee, H.J. (1991). Molecular cloning and characterization of genes expressed in shoot apical meristems. Plant Cell **3,** 359-370.

Mejia-Guerra, M.K., Li, W., Galeano, N.F., Vidal, M., Gray, J., Doseff, A.I., and Grotewold, E. (2015). Core Promoter Plasticity Between Maize Tissues and Genotypes Contrasts with Predominance of Sharp Transcription Initiation Sites. Plant Cell **27,** 3309-3320.

Millar, A.A., and Gubler, F. (2005). The Arabidopsis GAMYB-like genes, MYB33 and MYB65, are microRNA-regulated genes that redundantly facilitate anther development. The Plant cell **17,** 705-721.

Mo, A., Mukamel, E.A., Davis, F.P., Luo, C., Henry, G.L., Picard, S., Urich, M.A., Nery, J.R., Sejnowski, T.J., Lister, R., Eddy, S.R., Ecker, J.R., and Nathans, J. (2015). Epigenomic Signatures of Neuronal Diversity in the Mammalian Brain. Neuron **86,** 1369-1384.

Morton, T., Petricka, J., Corcoran, D.L., Li, S., Winter, C.M., Carda, A., Benfey, P.N., Ohler, U., and Megraw, M. (2014). Paired-end analysis of transcription start sites in Arabidopsis reveals plant-specific promoter signatures. Plant Cell **26,** 2746-2760.

Mumbach, M.R., Satpathy, A.T., Boyle, E.A., Dai, C., Gowen, B.G., Cho, S.W., Nguyen, M.L., Rubin, A.J., Granja, J.M., Kazane, K.R., Wei, Y., Nguyen, T., Greenside, P.G., Corces, M.R., Tycko, J., Simeonov, D.R., Suliman, N., Li, R., Xu, J., Flynn, R.A., Kundaje, A., Khavari, P.A., Marson, A., Corn, J.E., Quertermous, T., Greenleaf, W.J., and Chang, H.Y. (2017). Enhancer connectome in primary human cells identifies target genes of disease-associated DNA elements. Nat Genet **49,** 1602-1612.

Murashige, T., and Skoog, F. (1962). A Revised Medium for Rapid Growth and Bio Assays with Tobacco Tissue Cultures. Physiol Plantarum **15,** 473-497.

Nogales-Cadenas, R., Carmona-Saez, P., Vazquez, M., Vicente, C., Yang, X., Tirado, F., Carazo, J.M., and Pascual-Montano, A. (2009). GeneCodis: interpreting gene lists through enrichment analysis and integration of diverse biological information. Nucleic Acids Res **37,** W317-322.

O'Malley, R.C., Huang, S.C., Song, L., Lewsey, M.G., Bartlett, A., Nery, J.R., Galli, M., Gallavotti, A., and Ecker, J.R. (2016). Cistrome and Epicistrome Features Shape the Regulatory DNA Landscape. Cell **166,** 1598.

Oh, S.J., Song, S.I., Kim, Y.S., Jang, H.J., Kim, S.Y., Kim, M., Kim, Y.K., Nahm, B.H., and Kim, J.K. (2005). Arabidopsis CBF3/DREB1A and ABF3 in transgenic rice increased tolerance to abiotic stress without stunting growth. Plant Physiol **138,** 341-351.

Ong, C.T., and Corces, V.G. (2011). Enhancer function: new insights into the regulation of tissue-specific gene expression. Nat Rev Genet **12,** 283-293.

Oyama, T., Shimura, Y., and Okada, K. (1997). The Arabidopsis HY5 gene encodes a bZIP protein that regulates stimulus-induced development of root and hypocotyl. Genes & Development **11,** 2983-2995.

Pajoro, A., Madrigal, P., Muino, J.M., Matus, J.T., Jin, J., Mecchia, M.A., Debernardi, J.M., Palatnik, J.F., Balazadeh, S., Arif, M., O'Maoileidigh, D.S., Wellmer, F., Krajewski, P., Riechmann, J.L., Angenent, G.C., and Kaufmann, K. (2014). Dynamics of chromatin accessibility and gene regulation by MADS-domain transcription factors in flower development. Genome Biol **15,** R41.

Patel, R.V., Nahal, H.K., Breit, R., and Provart, N.J. (2012). BAR expressolog identification: expression profile similarity ranking of homologous genes in plant species. The Plant Journal **71,** 1038-1050.

Persak, H., and Pitzschke, A. (2014). Dominant repression by Arabidopsis transcription factor MYB44 causes oxidative damage and hypersensitivity to abiotic stress. International Journal of Molecular Sciences **15,** 2517-2537.

Piper, J., Elze, M.C., Cauchy, P., Cockerill, P.N., Bonifer, C., and Ott, S. (2013). Wellington: a novel method for the accurate identification of digital genomic footprints from DNase-seq data. Nucleic Acids Research **41,** e201-e201.

Piper, J., Assi, S.A., Cauchy, P., Ladroue, C., Cockerill, P.N., Bonifer, C., and Ott, S. (2015). Wellington-bootstrap: differential DNase-seq footprinting identifies cell-type determining transcription factors. BMC genomics **16,** 1000.

Ramirez, F., Ryan, D.P., Gruning, B., Bhardwaj, V., Kilpert, F., Richter, A.S., Heyne, S., Dundar, F., and Manke, T. (2016). deepTools2: a next generation web server for deep-sequencing data analysis. Nucleic Acids Res **44,** W160-165.

Reyes, J.L., and Chua, N.H. (2007). ABA induction of miR159 controls transcript levels of two MYB factors during Arabidopsis seed germination. Plant J **49,** 592-606.

Rodgers-Melnick, E., Vera, D.L., Bass, H.W., and Buckler, E.S. (2016). Open chromatin reveals the functional maize genome. Proceedings of the National Academy of Sciences of the United States of America **113,** E3177-E3184.

Ron, M., Kajala, K., Pauluzzi, G., Wang, D., Reynoso, M.A., Zumstein, K., Garcha, J., Winte, S., Masson, H., Inagaki, S., Federici, F., Sinha, N., Deal, R.B., Bailey-Serres, J., and Brady, S.M. (2014). Hairy root transformation using Agrobacterium rhizogenes as a tool for exploring cell type-specific gene expression and function using tomato as a model. Plant Physiol **166,** 455-469.

Ruzicka, D.R., Kandasamy, M.K., McKinney, E.C., Burgos-Rivera, B., and Meagher, R.B. (2007). The ancient subclasses of Arabidopsis Actin Depolymerizing Factor genes exhibit novel and differential expression. Plant J **52,** 460-472.

Salmon-Divon, M., Dvinge, H., Tammoja, K., and Bertone, P. (2010). PeakAnalyzer: genome-wide annotation of chromatin binding and modification loci. BMC Bioinformatics **11,** 415.

Scharer, C.D., Blalock, E.L., Barwick, B.G., Haines, R.R., Wei, C., Sanz, I., and Boss, J.M. (2016). ATAC-seq on biobanked specimens defines a unique chromatin accessibility structure in naïve SLE B cells. Nature Publishing Group **6,** 27030.

Schellmann, S., Schnittger, A., Kirik, V., Wada, T., Okada, K., Beermann, A., Thumfahrt, J., Jurgens, G., and Hulskamp, M. (2002). TRIPTYCHON and CAPRICE mediate lateral inhibition during trichome and root hair patterning in Arabidopsis. EMBO J **21,** 5036-5046.

Shim, J.S., Jung, C., Lee, S., Min, K., Lee, Y.W., Choi, Y., Lee, J.S., Song, J.T., Kim, J.K., and Choi, Y.D. (2013). AtMYB44 regulates WRKY70 expression and modulates antagonistic interaction between salicylic acid and jasmonic acid signaling. Plant J **73,** 483-495.

Shin, R., Burch, A.Y., Huppert, K.A., Tiwari, S.B., Murphy, A.S., Guilfoyle, T.J., and Schachtman, D.P. (2007). The Arabidopsis transcription factor MYB77 modulates auxin signal transduction. The Plant cell **19,** 2440-2453.

Song, L., Huang, S.s.C., Wise, A., Castanon, R., Nery, J.R., Chen, H., Watanabe, M., Thomas, J., Bar-Joseph, Z., and Ecker, J.R. (2016). A transcription factor hierarchy defines an environmental stress response network. Science (New York, NY) **354,** aag1550-aag1550.

Spitz, F., and Furlong, E.E. (2012). Transcription factors: from enhancer binding to developmental control. Nat Rev Genet **13,** 613-626.

Stadhouders, R., van den Heuvel, A., Kolovos, P., Jorna, R., Leslie, K., Grosveld, F., and Soler, E. (2012). Transcription regulation by distal enhancers: who's in the loop? Transcription **3,** 181-186.

Sullivan, A.M., Arsovski, A.A., Lempe, J., Bubb, K.L., Weirauch, M.T., Sabo, P.J., Sandstrom, R., Thurman, R.E., Neph, S., Reynolds, A.P., Stergachis, A.B., Vernot, B., Johnson, A.K., Haugen, E., Sullivan, S.T., Thompson, A., Neri III, F.V., Weaver, M., Diegel, M., Mnaimneh, S., Yang, A., Hughes, T.R., Nemhauser, J.L., Queitsch, C., and Stamatoyannopoulos, J.A. (2014). Mapping and Dynamics of Regulatory DNA and Transcription Factor Networks in Arabidopsis thaliana. CellReports **8,** 2015-2030.

Sung, M.H., Baek, S., and Hager, G.L. (2016). Genome-wide footprinting: ready for prime time? Nat Methods **13,** 222-228.

Tabas-Madrid, D., Nogales-Cadenas, R., and Pascual-Montano, A. (2012). GeneCodis3: a non-redundant and modular enrichment analysis tool for functional genomics. Nucleic Acids Res **40,** W478-483.

Thorvaldsdottir, H., Robinson, J.T., and Mesirov, J.P. (2013). Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. Brief Bioinform **14,** 178-192.

Thurman, R.E., Rynes, E., Humbert, R., Vierstra, J., Maurano, M.T., Haugen, E., Sheffield, N.C., Stergachis, A.B., Wang, H., Vernot, B., Garg, K., John, S., Sandstrom, R., Bates, D., Boatman, L., Canfield, T.K., Diegel, M., Dunn, D., Ebersol, A.K., Frum, T., Giste, E., Johnson, A.K., Johnson, E.M., Kutyavin, T., Lajoie, B., Lee, B.K., Lee, K., London, D., Lotakis, D., Neph, S., Neri, F., Nguyen, E.D., Qu, H., Reynolds, A.P., Roach, V., Safi, A., Sanchez, M.E., Sanyal, A., Shafer, A., Simon, J.M., Song, L., Vong, S., Weaver, M., Yan, Y., Zhang, Z., Zhang, Z., Lenhard, B., Tewari, M., Dorschner, M.O., Hansen, R.S., Navas, P.A., Stamatoyannopoulos, G., Iyer, V.R., Lieb, J.D., Sunyaev, S.R., Akey, J.M., Sabo, P.J., Kaul, R., Furey, T.S., Dekker, J., Crawford, G.E., and Stamatoyannopoulos, J.A. (2012). The accessible chromatin landscape of the human genome. Nature **489,** 75-82.

Tian, T., Liu, Y., Yan, H., You, Q., Yi, X., Du, Z., Xu, W., and Su, Z. (2017). agriGO v2.0: a GO analysis toolkit for the agricultural community, 2017 update. Nucleic Acids Res.

Tittarelli, A., Santiago, M., Morales, A., Meisel, L.A., and Silva, H. (2009). Isolation and functional characterization of cold-regulated promoters, by digitally identifying peach fruit cold-induced genes from a large EST dataset. BMC Plant Biol **9,** 121.

Vanneste, K., Baele, G., Maere, S., and Van de Peer, Y. (2014). Analysis of 41 plant genomes supports a wave of successful genome duplications in association with the Cretaceous-Paleogene boundary. Genome Res **24,** 1334-1347.

Vera, D.L., Madzima, T.F., Labonne, J.D., Alam, M.P., Hoffman, G.G., Girimurugan, S.B., Zhang, J., McGinnis, K.M., Dennis, J.H., and Bass, H.W. (2014). Differential nuclease sensitivity profiling of chromatin reveals biochemical footprints coupled to gene expression and functional DNA elements in maize. Plant Cell **26,** 3883-3893.

Vierstra, J., and Stamatoyannopoulos, J.A. (2016). Genomic footprinting. Nature Methods **13,** 213-221.

Wada, T., Kurata, T., Tominaga, R., Koshino-Kimura, Y., Tachibana, T., Goto, K., Marks, M.D., Shimura, Y., and Okada, K. (2002). Role of a positive regulator of root hair development, CAPRICE, in Arabidopsis root epidermal cell differentiation. Development **129,** 5409-5419.

Wang, N., Xu, H., Jiang, S., Zhang, Z., Lu, N., Qiu, H., Qu, C., Wang, Y., Wu, S., and Chen, X. (2017). MYB12 and MYB22 play essential roles in proanthocyanidin and flavonol synthesis in red-fleshed apple (Malus sieversii f. niedzwetzkyana). Plant J **90,** 276-292.

Wang, Z., Su, G., Li, M., Ke, Q., Kim, S.Y., Li, H., Huang, J., Xu, B., Deng, X.P., and Kwak, S.S. (2016). Overexpressing Arabidopsis ABF3 increases tolerance to multiple abiotic stresses and reduces leaf size in alfalfa. Plant Physiol Biochem **109,** 199-208.

Weber, B., Zicola, J., Oka, R., and Stam, M. (2016). Plant Enhancers: A Call for Discovery. Trends in plant science **21,** 974-987.

Weirauch, M.T., Yang, A., Albu, M., Cote, A., Montenegro-Montero, A., Drewe, P., Najafabadi, H.S., Lambert, S.A., Mann, I., Cook, K., Zheng, H., Goity, A., van Bakel, H., Lozano, J.C., Galli, M., Lewsey, M., Huang, E., Mukherjee, T., Chen, X., Reece-Hoyes, J.S., Govindarajan, S., Shaulsky, G., Walhout, A.J.M., Bouget, F.Y., Ratsch, G., Larrondo, L.F., Ecker, J.R., and Hughes, T.R. (2014). Determination and inference of eukaryotic transcription factor sequence specificity. Cell **158,** 1431-1443.

Wilkins, O., Hafemeister, C., Plessis, A., Holloway-Phillips, M.M., Pham, G.M., Nicotra, A.B., Gregorio, G.B., Jagadish, S.V., Septiningsih, E.M., Bonneau, R., and Purugganan, M. (2016). EGRINs (Environmental Gene Regulatory Influence Networks) in Rice That Function in the Response to Water Deficit, High Temperature, and Agricultural Environments. Plant Cell **28,** 2365-2384.

Xu, Y., and Du, J. (2014). Young but not relatively old retrotransposons are preferentially located in gene-rich euchromatic regions in tomato (Solanum lycopersicum) plants. Plant J **80,** 582-591.

Yanhui, C., Xiaoyuan, Y., Kun, H., Meihua, L., Jigang, L., Zhaofeng, G., Zhiqiang, L., Yunfei, Z., Xiaoxiao, W., Xiaoming, Q., Yunping, S., Li, Z., Xiaohui, D., Jingchu, L., Xing-Wang, D., Zhangliang, C., Hongya, G., and Li-Jia, Q. (2006). The MYB transcription factor superfamily of Arabidopsis: expression analysis and phylogenetic comparison with the rice MYB family. Plant Mol Biol **60,** 107-124.

Zhang, W., Zhang, T., Wu, Y., and Jiang, J. (2012a). Genome-wide identification of regulatory DNA elements and protein-binding footprints using signatures of open chromatin in Arabidopsis. Plant Cell **24,** 2719-2731.

Zhang, W., Wu, Y., Schnable, J.C., Zeng, Z., Freeling, M., Crawford, G.E., and Jiang, J. (2012b). High-resolution mapping of open chromatin in the rice genome. Genome Research **22,** 151-162.

Zhao, Y., Xing, L., Wang, X., Hou, Y.J., Gao, J., Wang, P., Duan, C.G., Zhu, X., and Zhu, J.K. (2014). The ABA receptor PYL8 promotes lateral root growth by enhancing MYB77-dependent transcription of auxin-responsive genes. Sci Signal **7,** ra53.

Zhu, B., Zhang, W., Zhang, T., Liu, B., and Jiang, J. (2015). Genome-Wide Prediction and Validation of Intergenic Enhancers in Arabidopsis Using Open Chromatin Signatures. Plant Cell **27,** 2415-2426.

CHAPTER 4: DIFFERENCES IN DIRECTIONALITY OF RNA POLYMERASE INITIATION UNDERLIE EPIGENOME DIFFERENCES BETWEEN PLANTS AND ANIMALS

Kelsey A. Maher[1,2], Dongxue Wang[3], and Roger B. Deal[1*]

[1]Department of Biology

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

[3]Department of Biochemistry

Emory University, Atlanta, GA 30322 USA

[*]Correspondence: Roger B. Deal; roger.deal@emory.edu

This manuscript is in preparation for submission to G3: Genes, Genomes, Genetics

Kelsey A. Maher generated the *Arabidopsis* ATAC-seq dataset and performed all the data processing, analysis, and figure generation. She also wrote and prepared the manuscript.

ABSTRACT

Transcriptional regulation is a universal mechanism for a wide array of biological processes, and is driven in large part by genetic enhancer elements. These regulatory elements have been well-studied in animal species, yet their plant counterparts remain poorly characterized. While high-throughput profiling of animal genomes has yielded great success in identifying genetic enhancers through secondary characteristics – flanking histone posttranslational modifications (PTMs), chromatin accessibility, and the production of enhancer RNAs (eRNAs) – it is an active line of investigation as to whether these secondary characteristics can be used in a similar way to locate regulatory regions of plant genomes. Here, we compare the enrichment of the four histone PTMs most commonly associated with animal enhancers – H3K27ac, H3K27me3, H3K4me1, and H3K4me3 – between *Drosophila melanogaster*, *Homo sapiens*, and *Arabidopsis thaliana* genomes. Regions of accessible chromatin were identified through ATAC-seq or DNase-seq, and were analyzed as putative enhancer regions. Additionally, as it has been shown that

enhancer activity varies widely from cell type to cell type, matched, single-cell type datasets were used for each species whenever available. Through the intersection of these data it becomes clear that there are distinct differences between the epigenetic makeup of plant and animal genomes. While these four histone PTMs are present at transcription start sites (TSSs) in all three of the species investigated, *A. thaliana* showed a marked depletion of these modifications upstream of the TSS, while the animal species showed bimodal enrichment. The plant histone PTM pattern is consistent with the pattern observed at unidirectional promoters, which was further supported by GRO-seq data. When intergenic regions of accessible chromatin were examined – putative enhancer regions – the plant epigenomes showed a one-sided, rather than bimodal, enrichment of all four of the histone PTMs. However, these sites retain the ability to produce eRNAs, suggesting that they are likely functionally active enhancer elements. While it is known that animal promoters and enhancers have bidirectional transcription, this analysis revealed that plant promoters and enhancers have a distinct pattern, and only exhibit histone PTM deposition and transcription in the sense direction. While further investigation is merited, this may speak to a fundamental difference between the transcriptional machinery of plant and animal kingdoms.

INTRODUCTION

In order to survive, all species must control gene expression in a manner that is both cell type-specific and adaptive to changing external cues. To meet this challenge, organisms use *cis*-acting sequences of DNA to regulate the activity of promoters in order to modulate the transcriptional output of nearby genes. Many subcategories of these regulatory DNA sequences have been uncovered, including silencers and insulators, but perhaps the best characterized is the enhancer element.

Enhancers are a highly conserved type of genetic control mechanism, and have been found in a diverse array of organisms, from eukaryotes (Schwaiger, Schonauer et al. 2014, Villar, Berthelot et al. 2015, Zhu, Zhang et al. 2015, Weber, Zicola et al. 2016) to bacteria (Xu and Hoover 2001) and viruses (Berg, Popovic et al. 1984). On a molecular scale, genetic enhancers consist of DNA sequences ranging between

tens to hundreds of base pairs in length. These sequences are comprised of a modular collection of transcription factor binding motifs which in turn act as an assembly platform for *trans*-acting factors (Lee and Young 2000, Spitz and Furlong 2012). 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, RNA Polymerase II, nucleosome remodelers, and histone modifying proteins such as CPB/p300 (Vernimmen and Bickmore 2015). Ultimately, the *cis*-regulatory element and its associated *trans*-factors will maneuver to interact with a target promoter, enriching the local microenvironment for activating TFs and assembled transcriptional machinery in order to promote the transcription of the target gene. In this manner, enhancers act as a mechanism to integrate a myriad of cellular signals into meaningful transcriptional output.

Enhancers provide a crucial and necessary function to the cell, but these elements have historically presented an enormous challenge to study. Promoters can be readily identified by a suite of characteristics, from conserved sequence motifs (TATA boxes and CpG islands) to their proximal location upstream of gene bodies (Kim and Shiekhattar 2015). Enhancers, in contrast, have startling few commonalities between elements. Beyond the DNA-binding motifs of individual TFs, this class of *cis*-regulatory sequences lacks any sort of overarching sequence conservation (Villar, Berthelot et al. 2015). Enhancers have been found to be able to act in both 'forward' and 'reverse' orientations with regard to their target promoter, and functional elements have been found to be abundant in both the genic and intergenic regions of the genome. Additionally, while evidence suggests that enhancers preferentially regulate the most proximal promoter (Heintzman, Hon et al. 2009, Anders and Huber 2010, Creyghton, Cheng et al. 2010), other elements have been characterized up to tens of thousands of base pairs away from their target promoters, or even on entirely separate chromosomes (Lettice, Heaney et al. 2003, Kleinjan and van Heyningen 2005, Ong and Corces 2011). Taken together, this remarkably permissive profile leaves very little concrete criteria on which to positively identify an enhancer element in a discriminating manner, making the study of these elements uniquely challenging.

Significant headway has been made to overcome these hurdles with the advent of next-generation sequencing assays. Techniques that probe the relative accessibility of chromatin to *trans*-acting factors, including DNase-seq (Keene, Corces et al. 1981, McGhee, Wood et al. 1981, Boyle, Davis et al. 2008) and ATAC-seq (Buenrostro, Giresi et al. 2013, Buenrostro, Wu et al. 2015), reveal that enhancers and other regulatory regions of DNA preferentially exist in areas of accessible chromatin. Furthermore, through the use of genome-wide chromatin immunoprecipitation with high-throughput sequencing (ChIP-seq), a variety of histone posttranslational modifications (PTMs) have been found to be associated with the well-positioned nucleosomes that flank enhancers (Wang, Zang et al. 2008, Hawkins, Hon et al. 2010, Ernst, Kheradpour et al. 2011, Zentner, Tesar et al. 2011, Bonn, Zinzen et al. 2012). Among these studies, a single set of marks has emerged as the most highly conserved across enhancers in a variety of cell types and species: H3K27ac, H3K27me3, H3K4me1, and H3K4me3. As several of these modifications have also been reported to have overlap with promoters, a higher ratio of H3K4me1/H3K4me3 enrichment has been used as a guideline to distinguish enhancers from their gene-proximal counterparts (Heintzman, Stuart et al. 2007, Heintzman, Hon et al. 2009, Kim and Shiekhattar 2015).

The next-generation sequencing approaches of the ENCODE Project (2004) have catapulted the characterization of *cis*-activating regulatory elements in humans and animal models. Unfortunately, the state-of-the-art approach utilized in plant systems has lagged decades behind. Much of the research that exists on plant enhancers has relied largely on labor-intensive reporter cloning assays, greatly slowing down the progress of the field. Until recently, only a handful of genuine enhancer elements have been functionally characterized across the entire Plantae kingdom (Zhu, Zhang et al. 2015, Weber, Zicola et al. 2016, Yan, Chen et al. 2019). With such a small population of elements to analyze, it remained an open question whether or not the unique set of enhancer characteristics conserved across animal species is similarly conserved in plant species. Significant advances have been made using DNase I-hypersensitive sites (DHSs) as a marker to identify putative enhancers genome-wide in select plant species (Zhang, Wu et al. 2012, Zhang, Zhang et al. 2012, Pajoro, Madrigal et al. 2014, Zhu, Zhang et al. 2015, Oka, Zicola et al. 2017, Wang, Tu et al. 2017). Building on this success, accessible chromatin regions (ACRs) identified by

ATAC-seq were assayed for a variety of histone PTMs in thirteen angiosperm species (Lu, Marand et al. 2019). However, all of these studies rely on whole tissue or whole organisms as input. In line with their role as drivers of embryonic development and cell specification programs (Chatterjee and Ahituv 2017), the activity of enhancers varies drastically in regard to individual cell type (Ernst and Kellis 2010, Rada-Iglesias, Bajpai et al. 2011, Zentner, Tesar et al. 2011). As such, it is ideal to compare sequencing datasets drawn from the same cell type. Studies which examine chromatin signatures across a whole tissue or organism risk muddying enhancer activity state signals by averaging enrichment trends together across several cell types (Creyghton, Cheng et al. 2010, Rada-Iglesias, Bajpai et al. 2011). This effect would dampen regions of extreme enrichment or depletion, making genuine elements more difficult to distinguish from background noise.

In this study, we generate ChIP-seq datasets with single-cell type specificity for the highly conserved set of animal enhancer histone PTMs H3K27ac, H3K27me3, H3K4me1, and H3K4me3, in the model plant *Arabidopsis thaliana*. Combined with single-cell type ATAC-seq data previously generated by our lab and available GRO-seq data, we are able to probe the extent of conserved enhancer characteristics on multiple levels. When compared with available single-cell type datasets for *Homo sapiens* and *Drosophila melanogaster*, we find that the bimodal pattern of histone PTMs at animal enhancers is not conserved in *Arabidopsis*. Instead, accessible chromatin regions are exclusively flanked on one side or the other by the characteristic PTMs. Transcription preferentially proceeds unidirectionally in *Arabidopsis*, leading to histone PTM deposition in the sense direction alone, both at TSSs and at intergenic hyperaccessible sites. While it is known that animal promoters and enhancers are transcribed bidirectionally, this analysis revealed that plant promoters and enhancers have a distinct pattern, and may speak to a fundamental difference in RNA polymerase II initiation between the plant and animal kingdoms.

RESULTS

PROMOTER TRANSCRIPTION IS BIDIRECTIONAL IN ANIMAL MODELS AND UNIDIRECTIONAL IN *ARABIDOPSIS*

Using ATAC-seq coupled with the cell type-specific nuclei purification technology of INTACT, we identified 30,962 discrete accessible sites within the epigenome of *Arabidopsis thaliana* non-hair root epidermal cells from 10-day old seedlings. Also using INTACT, we generated cell type-specific ChIP-seq datasets for the conserved animal enhancer histone modifications H3K27ac, H3K27me3, H3K4me1, and H3K4me3, and plotted them on the accessible chromatin regions (ACRs). The matched cell type data in this analysis gives us the critical ability to examine the relationship between chromatin accessibility and histone modifications in the plant epigenome without signal interference from multiple cell types.

Increasingly, studies have revealed a trend that enhancer and promoter elements are highly similar in their structure and functionality (Ernst, Kheradpour et al. 2011, Kim and Shiekhattar 2015). As enhancer elements have yet to be rigorously defined in plant species, we began our investigation by comparing single-cell type histone modification enrichment and chromatin accessibility at genic sites in *A. thaliana*, *H. sapiens*, and *D. melanogaster*. The superior annotation of these basal regulatory elements offers strong grounds for direct comparison between the eukaryotic kingdoms, and offers insights into trends in global transcriptional regulation that may illuminate similar mechanisms in putative enhancers. **Figure 4.1A** shows metaplots of the average single-cell type ChIP-seq signal for H3K27ac, H3K27me3, H3K4me1, and H3K4me3 enrichment and chromatin accessibility across gene bodies in each of the three species of interest.

Even at this global scale, broad similarities are apparent in the pattern of chromatin accessibility (single-cell type DNase-seq for *H. sapiens* and *D. melanogaster*, single-cell type ATAC-seq for *A. thaliana*) across the gene body. The region of maximum hyperaccessibility is restricted to a narrow peak 100-250 bp directly upstream of the TSS, and is primarily due to the activity of RNA Polymerase II (RNAPII). This holoenzyme acts as a chromatin remodeler as it progresses along the transcribed region in its path (Core, Waterfall et al. 2008), sliding and evicting nucleosomes to allow other *trans*-acting factors – including DNase and transposase – access to the underlying DNA. In spite of this fundamental similarity, 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 *H. sapiens* and *D. melanogaster* metaplots. This pattern of enrichment is attributed to the divergent nature of animal promoters and their proclivity to produce transcripts from a single TSS in both the sense and antisense directions (Trinklein, Aldred et al. 2004, Kim, Barrera et al. 2005, Barski, Cuddapah et al. 2007, Guenther, Levine et al. 2007, Core, Waterfall et al. 2008). Not only does RNAPII remodel nucleosomes as it progresses (Core, Waterfall et al. 2008), the active, phosphorylated version of RNAPII acts as a binding platform for histone modifying complexes, such as MLL3 and MLL4 in mammals. These complexes in turn deposit modifications, such as H3K4me1/2, on the underlying histones successively through multiple rounds of elongation (Kaikkonen, Spann et al. 2013). As such, the process of transcription itself is responsible for maintaining the accessible chromatin structure found at the site of transcriptional initiation, as well as for the surrounding deposition of the characteristic set of histone modifications (Seila, Core et al. 2009). This process leads to the enrichment of histone PTMs both upstream and downstream of the accessible TSS region in animals, as is shown in the above metaplots.

In contrast, a unique pattern is seen at the TSS of the *Arabidopsis* metaplot. The histone modification ChIP-seq signal is most abundant at the 5' end of gene bodies, with the signal upstream of transcription start sites reduced to near background levels. In light of the mechanisms responsible for generating and maintaining the bimodal enrichment of histone modifications around animal TSSs, the absence of this pattern strongly suggests that – distinct from the more promiscuous transcriptional process in animals – transcription in plants may proceed exclusively in the sense orientation, accounting for the sole downstream presence of histone marks. In order to examine this possibility more closely, we analyzed publicly available Global Run-On sequencing (GRO-seq) (Core, Waterfall et al. 2008, Melgar, Collins et al. 2011, Hah, Murakami et al. 2013, Gardini 2017) data from *H. sapiens*, *D. melanogaster*, and *A. thaliana*. We separated all protein-coding genes across these genomes based on their strandedness, plotting all plus strand genes (**Figure 4.1B**) and minus strand genes (**Figure 4.1C**) separately. Within gene bodies, the 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. It is worth highlighting that just upstream of the TSS in *H. sapiens* and *D. melanogaster* genomes, transcripts running opposite of the genic direction are produced, as is typical of divergent transcription at promoters (Kapranov, Cheng et al. 2007, Core, Waterfall et al. 2008, Seila, Calabrese et al. 2008) and enhancers (Kim, Hemberg et al. 2010, Hah, Murakami et al. 2013, Shlyueva, Stampfel et al. 2014). This absence of upstream signal in *Arabidopsis* of transcripts of either strand indicates that transcription is truly one-directional in this organism. This directly matches what we observed from the enrichment of histone PTM signal, and further supports that histone PTM enrichment directly reflects transcriptional output. In addition to emerging in the metaplot, this bimodal/unimodal enrichment pattern of histone modifications is recapitulated at TSSs across the genomes of these species (**Supplemental Figure 4.1**).

PROXIMAL INTERGENIC ACCESSIBLE CHROMATIN REGIONS ARE FLANKED ON A SINGLE SIDE BY CHARACTERISTIC ENHANCER MARKS IN *ARABIDOPSIS*

The goal of our investigation was to compare the epigenetic signature of enhancers in *Arabidopsis* with what has been previously established in animal models. However, enhancer elements have yet to be broadly annotated in plant genomes, making direct element-to-element comparisons highly limited. To overcome this obstacle, we examined regions of chromatin hyperaccessibility as defined by particular susceptibility to activity by DNase I or transposase Tn5. Enhancers have been shown to preferentially reside in regions of accessible chromatin (Tsompana and Buck 2014, Jiang 2015), and hyperaccessible sites have been used previously as markers of putative regulatory elements (Bell, Tiwari et al. 2011). We began by mapping the four canonical enhancer histone modifications onto non-genic accessible chromatin regions (ACRs) detected by ATAC-seq in *Arabidopsis* (**Figure 4.2**). Chromatin accessibility is a critical feature of regulatory elements, allowing them to associate with *trans*-acting factors such as transcription factors and chromatin remodelers. We began by plotting the average signal across all proximal intergenic accessible chromatin regions (2 kb to 100 bp upstream of an annotated TSS/100 bp to 1 kb downstream of an annotated TES). Previous investigations into plant enhancers restricted the search to regions of accessibility that are

at least 2 kb upstream of a transcription start site (Zhu, Zhang et al. 2015). This is biologically reasonable as many enhancer elements characterized in animal species have been shown to be able to act from several kilobases away from their target promoters (Lettice, Heaney et al. 2003, Kleinjan and van Heyningen 2005, Ong and Corces 2011), and such a cutoff would eliminate many of the false positives from promoter elements. However, while some enhancers are distant from their targets, it has been shown that the majority of elements regulate their most proximal gene (Mendenhall, Williamson et al. 2013, Ghavi-Helm, Klein et al. 2014, Kvon, Kazmar et al. 2014). Particularly in plants, the majority of non-genic ACRs in the *Arabidopsis*, rice, tomato, and *Medicago* genomes fall within 2 kb of the TSS (Maher, Bajic et al. 2018). Nearly 60% of the ACRs in the *Arabidopsis* accessibility data fall within the proximal intergenic region (+2 kb-100 bp upstream of TSS/-100 bp-1 kb downstream of TES) (**Supplementary Figure 4.2**). ACRs have also been shown to cluster close to the TSS in a variety of other angiosperm species, both monocots and dicots (Lu, Marand et al. 2019). These findings have been mirrored by the small interacting regions (kilobase-sized) uncovered by *Arabidopsis* Hi-C chromosome conformation assays (Feng, Cokus et al. 2014, Wang, Liu et al. 2015). As such, there is compelling evidence to suggest that there are biologically relevant regulatory elements near gene bodies.

The metaplot of histone modification enrichment at proximal intergenic accessible chromatin regions (**Figure 4.2**) shows an area of expanded, localized depletion spanning the width of the accessible peak (median peak size 264 bp, +/- 192 bp). Directly flanking the ACR on each side are symmetrical peaks of enrichment for H3K27ac and H3K4me3, which is in line with what has been previously reported in animal studies. While regulatory elements contain '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 (Schones, Cui et al. 2008, Henikoff, Henikoff et al. 2009, Jin, Zang et al. 2009). As such, this enrichment pattern observed in *A. thaliana* is not dissimilar from the pattern observed at proximal intergenic ACRs in *H. sapiens* and *D. melanogaster*. The ACRs (median peak size 164 bp +/- 245 bp and 129 bp +/- 232 bp respectively) are flanked on both sides by a pronounced enrichment for H3K27ac, H3K4me1, and H3K4me3. 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 it transitions from initiation to elongation (Kaikkonen, Spann et al. 2013), as is reflected in the production of nascent transcripts surrounding the accessible chromatin region (**Figure 4.2**). The notable histone deposition present in the center of the animal metaplots can also be explained by the averaging of overlapping signal of bidirectional antisense transcription, which produces an artificially large signal associated with the ACR peak.

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. In **Figure 4.2**, intergenic accessible chromatin regions across the human, *Drosophila*, and *Arabidopsis* genomes are mapped in a heatmap, grouped into ten subpopulations via k-means clustering. In addition to H3K27ac, H3K27me3, H3K4me1, H3K4me3, and chromatin accessibility data, we have also generated single-cell type ChIP-seq datasets for histones H3 and H2A.Z in the *Arabidopsis* root epidermal non-hair cell. H2A.Z is a histone variant associated with the flanking regions of active enhancers (Jin, Zang et al. 2009), while canonical histone H3 is a fundamental component of histone octamers. In addition to histone data, we have included publicly available ChIP-seq data for ABF3. ABF3 is a transcription factor that is highly expressed in the root tip of *Arabidopsis* (Maher, Bajic et al. 2018). Sites of chromatin hyperaccessibility across the *Arabidopsis* non-hair cell genome are precisely mirrored by ABF3 ChIP-seq signal enrichment. Enhancers and other *cis*-regulatory elements function as, first and foremost, a binding platform for *trans*-acting factors such as the TF ABF3. While the overlap in location of ACRs and TF binding sites is not in and of itself conclusive that these sequences are enhancers, it does support that these regions meet one of the crucial criteria of the regulatory element class.

A prominent pattern that emerges from the subclusters of **Figure 4.2** is an enrichment for the posttranslational histone modifications on one side of the accessible peak or the other, but not both. All four of the histone marks, as well as the two histone tracks, show overlapping enrichment upstream or downstream of their respective accessible peaks but never surround the site on both sides simultaneously.

A prime example of this is Cluster 6 (**C6**). High intensity signal can be seen in the center of the window in the ATAC-seq heatmap, indicating a pronounced region of high chromatin accessibility. Directly 'upstream' of this region in the H3K27ac, H3K27me3, H3K4me1, and H3K4me3 heatmaps a strong enrichment of ChIP-seq signal can be seen in each one of the plots. 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 (Shilatifard 2006, Li, Carey et al. 2007). Here, we can observe corresponding enrichments of H3K4me3 proximal to the ACR, with H3K4me1 signal presenting more distally. Distinct from most studied eukaryotes, however, our *Arabidopsis* data show dual enrichment for H3K27ac and H3K27me3 at the same loci. While these modifications are considered to be mutually exclusive in animal models, this simultaneous enrichment has been documented in previous *Arabidopsis* chromatin accessibility studies (Zhu, Zhang et al. 2015). 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 the rapid exchange of alternately modified histones/nucleosomes, is not yet clear. Hi-C experiments in *Arabidopsis* also uncover H3K27me3 as the major histone marker of interacting genomic regions (Feng, Cokus et al. 2014, Wang, Liu et al. 2015), suggesting that this modification may belie unique transcriptional regulatory machinery that exists within plants. This is further supported by a study done on *Arabidopsis* stomatal guard cells where modulation and redistribution of the levels of H3K27me3 in that cell type resulted in regulated developmental reprogramming (Lee, Wengier et al. 2019).

While the breadth and intensity of the signal varies between the histone marks, depending on the degree and spread of enrichment for that particular modification, the pattern remains consistent in that it is confined to the region adjacent to the ACR as defined by the ATAC-seq data. Additionally, this same adjacent region is also characterized by strong GRO-seq signal in the corresponding negative strand, in line with the evidence supporting the deposition of these PTMs during transcription itself (Kaikkonen, Spann et al. 2013). The pattern of signal enrichment observed for the histone posttranslational modifications is seen again for histones H2A.Z and H3 in Cluster 6, suggesting that these nucleosomes may be particularly

well-positioned (Schones, Cui et al. 2008, Henikoff, Henikoff et al. 2009, Jin, Zang et al. 2009). In contrast, Clusters 4, 5, and 8 (**C4, C5, C8**) shows marked enrichment on the opposite side of the prominent accessible chromatin region in the modification and histone heatmaps, and a corresponding increase in signal in the same region of the positive-strand GRO-seq heatmap.

The heatmap for *H. sapiens* displays consistent enrichment patterns as well (**Figure 4.2**). Many clusters show the staggered H3K4me3 and H3K4me1 enrichment of bidirectional transcription, paired with pronounced GRO-seq signal (**Clusters C1, C3, C4, C5, C7, C8, C10**). Many of the loci within these clusters showing moderate H3K27ac signal, as is typical of 'active' enhancers (Creyghton, Cheng et al. 2010, Rada-Iglesias, Bajpai et al. 2011). Because many of the modifications between enhancers and promoters have been reported to overlap, a higher ratio of H3K4me1/H3K4me3 enrichment has been used as a guideline to distinguish enhancers from their gene-proximal counterparts (Heintzman, Stuart et al. 2007, Heintzman, Hon et al. 2009, Kim and Shiekhattar 2015). Other loci clusters show enrichment for H3K27me3/H3K4me1 characteristic of 'poised/inactive' enhancers (Rada-Iglesias, Bajpai et al. 2011) (**Cluster C2**), and H3K4me1 alone, characteristic of 'intermediate' enhancers (Creyghton, Cheng et al. 2010) (**Cluster C9**).

A noteworthy difference from the *Arabidopsis* GRO-seq data is that active chromatin regions produce both positive-strand and negative-strand transcripts from a single locus, albeit with a bias for one direction over the other, as is typical of divergent promoters (Kapranov, Cheng et al. 2007, Core, Waterfall et al. 2008, Seila, Calabrese et al. 2008) and enhancers (Kim, Hemberg et al. 2010, Hah, Murakami et al. 2013, Shlyueva, Stampfel et al. 2014). Not every cluster is transcriptionally active however; regions enriched in H3K27me3 are transcriptionally repressed/silenced (**C6**). This phenomenon is due to the role of H3K27me3 in recruiting the transcription-suppressing Polycomb protein complex (Grossniklaus and Paro 2014), and can be observed in the *Drosophila* heatmap as well (**Clusters C7, C10**). Similar to the activity states detailed for humans, *Drosophila* enhancers have been described as falling into two major categories based on their associated histone PTMs, with developmental enhancers being enriched for H3K4me1 (**Clusters C2, C6, C8**), and housekeeping enhancers being enriched for H3K4me3 (Cubenas-Potts, Rowley

et al. 2017) (**Clusters C4, C5, C9**). The latter category also shows evidence of active divergent transcription (**Clusters C4, C5, C9**), as could be expected from housekeeping regulatory elements in terminally differentiated *Drosophila* S2 cells.

MANY DISTAL REGIONS ARE ACCESSIBLE, FEW ARE TRANSCRIPTIONALLY ACTIVE

The regions of the genome examined so far represents 89.8% of the accessible chromatin regions called in *Arabidopsis* (**Supplemental Figure 4.2**). While this proportion likely covers the majority of the regulatory sites in *Arabidopsis*, it also comes with a notable disadvantage. As genes and regulatory elements alike become activated and repressed, their epigenetic signatures are constantly being overwritten by the activity of *trans*-acting machinery coordinating the activity of the nearby *cis*-elements. Because distal regions are far removed from other major annotated genomic features, they have the highest chance of displaying unadulterated epigenetic enhancer signal. The average distance of a *Drosophila* enhancer to its target TSS is 10 kb (Kvon, Kazmar et al. 2014), while human enhancers average 130 kb from the promoters they regulate (Mumbach, Satpathy et al. 2017), suggesting that there are abundant regulatory elements in these genomes that are sufficiently far away to avoid having their characteristic marks being epigenetically overwritten. Therefore, due to their more protected status, distal accessible chromatin sites have represented the primary focus of high-throughput characterizations of plant enhancers (Zhu, Zhang et al. 2015, Oka, Zicola et al. 2017, Yan, Chen et al. 2019).

As might be expected, the metaplots of human and *Drosophila* (**Figure 4.3**) show varying levels of H3K27ac, H3K4me1, and H3K4me3 enrichment, reflective of an amalgamation of the multiple enhancer activity states described in these species. In stark contrast, however, the metaplot for *Arabidopsis* histone modification ChIP-seq data shows no distinction in enrichment of the distal sites from genomic background noise. Despite this prominent lack of canonical histone PTMs at *Arabidopsis* distal ACRs, there is notable GRO-seq signal localizing to the borders of these regions. While the degree of enrichment pales in comparison to the clear transcriptional output of *Drosophila* distal sites, the signal intensity of these transcripts are on par with those produced at human distal ACRs. Production of enhancer RNAs (eRNAs)

is a hallmark of animal enhancer elements (Lai and Shiekhattar 2014), and long noncoding RNAs (lncRNAs) have been characterized previously in *A. thaliana* that match the description of putative eRNAs (Liu, Jung et al. 2012, Wang, Chung et al. 2014, Yan, Chen et al. 2019). While the metaplot of distal hyperaccessible sites in *A. thaliana* does not resemble the characteristic profile of animal enhancers based on the criteria of histone mark enrichment, there is evidence that other criteria – particularly the modest production of putative eRNAs – are conserved.

The ten-cluster heatmap reveals that vast majority of the distal accessible chromatin regions in this cell type for *Arabidopsis* have little to no histone PTM enrichment or transcriptional activity. This could be expected in such a terminated differentiated cell type as the root epidermal non-hair cell. Previous studies (Maher, Bajic et al. 2018) have indicated that chromatin accessibility for the large part does not change from cell type to cell type in *Arabidopsis*. As such, even if a regulatory element is not in use in a particular cellular lineage, it is likely to remain accessible. The clusters that are transcriptionally active (**C5, C6**) show GRO-seq and ChIP-seq signal deposited on a single flanking side of the ACR, much in the pattern observed at gene bodies (**Figure 4.1**). However, these distal regions were chosen to be at least >2 kb upstream of a TSS, and >1 kb downstream of a TES, far outside the domain of genic transcription. The nascent transcripts present in these clusters are likely to be the long noncoding RNAs (lncRNAs), which have been previously likened to eRNAs in *Arabidopsis* (Liu, Jung et al. 2012, Wang, Chung et al. 2014, Yan, Chen et al. 2019). These lncRNAs/putative eRNAs are more prominent than comparable eRNA species generated at human or *Drosophila* distal accessible sites whose surrounding chromatin environment matches that of 'active' enhancer elements. These transcripts represent an intriguing phenomenon of putative enhancers in *Arabidopsis*, though their exact nature and function requires further investigation.

DISCUSSION

Enhancer elements have been historically challenging to study due to their remarkably permissive genetic and epigenetic profiles. Nowhere is this more apparent than in plant species, where investigations

have relied mainly on low-throughput reporter assays and chromatin accessibility profiling of whole genomes. These approaches are unable to account for the dynamic epigenetic activity states enhancers can exhibit from cell type to cell type, in line with their role as drivers of embryonic development and cell specification programs (Chatterjee and Ahituv 2017), and risk blurring data trends across heterogenous tissue samples. Our study provides the ChIP-seq datasets with single-cell type specificity for the highly conserved set of animal enhancer histone PTMs H3K27ac, H3K27me3, H3K4me1, and H3K4me3 in the model plant *Arabidopsis thaliana*. When examined alongside single-cell type ATAC-seq data previously generated by our lab and available GRO-seq data, critical differences become apparent in relation to transcriptional regulation in model animal species. Histone PTM enrichment at the 5' end of genes in *H. sapiens* and *D. melanogaster* shows a distinct bimodal distribution pattern around the transcription start site (TSS), while the enrichment pattern found around *A. thaliana* TSSs is noticeably missing the upstream mode. This bimodal pattern has been shown to be indicative of divergent transcription at the promoter (Kapranov, Cheng et al. 2007, Core, Waterfall et al. 2008, Seila, Calabrese et al. 2008), strongly implicating that in *Arabidopsis* transcription is more tightly regulated at the level of initiation. Further support for this interpretation is gained when DNase-hypersensitive sites and transposase-hypersensitive sites are examined across the genome. Putative regulatory regions are strongly correlated with regions of increased chromatin accessibility (Bell, Tiwari et al. 2011), an association that is well-supported by the histone PTM ChIP-seq data in animals.

In animal datasets these accessible chromatin sites match up well with the hallmark enhancer histone PTM pattern: surrounded by histones with high levels H3K27ac and H3K4me1, with relatively less H3K4me3 and H3K27me3. While initial metaplot results suggest a similarity in *Arabidopsis* accessible chromatin regions, clustered heatmaps reveal that this pattern is not observed genome-wide. Rather, the hallmark histone PTMs are found to be notably depleted at the ACR, and instead preferentially flank the region either upstream or downstream, but never both. While this trend challenges the current dogma about enhancer secondary characteristics, these accessible chromatin regions in *Arabidopsis* were found to fall in line with several other observed traits, including the ability to bind transcription factors and produce

transcriptional products, denoted as enhancer RNAs (eRNAs). The GRO-seq data reveals that these putative eRNAs at *Arabidopsis* accessible chromatin regions are long and unidirectional. Additionally, when the *Arabidopsis* ChIP-seq data were overlaid with the GRO-seq data, it was found that this one-sided flanking of the histone PTMs mapped directly with the production of transcripts.

Overall, the results of this investigation indicate that genuine differences exist between the plant and animal kingdom on the transcriptional level. While the elongation of transcripts in the sense direction appears to be preferred across all eukaryotes, the results of this study suggest that this direction is preferred with exclusivity in transcriptional initiation in plants, while animal transcription initiation is more promiscuous. This tendency is not element-specific and appears to take place at promoters and enhancers alike. This crucial, if subtle, difference radically shifts the mechanism by which these elements can be searched for. So much success has been garnered via the ENCODE project through large-scale sequencing projects, facilitated by the conservation of a unique histone PTM signature found on surrounding nucleosomes. The unidirectional nature of transcription found within plant species abolishes this trend, as these regions can only be flanked on one side by these histone PTMs. With the ability to pinpoint the precise location of a nearby regulatory element greatly reduced, these results describe a need for new criteria to identify these elements on a genome-wide scale.

STARR-seq is a high-throughput enhancer identification assay that has been used with great success in animal species (Arnold, Gerlach et al. 2013, Arnold, Gerlach et al. 2014, Shlyueva, Stampfel et al. 2014, Muerdter, Boryn et al. 2015, Cubenas-Potts, Rowley et al. 2017). Because this assay relies on an enhancer's fundamental ability to stimulate a target promoter to increase transcriptional output, it is able to identify functional elements without relying on secondary criteria, such as histone modifications. As our results point to the epigenetic enhancer signature being quite distinct from that which has been described in the literature for animal enhancers, this technique would be ideal for enhancer discovery in the plant genome, as it would uncouple positive identifications from much of the preconceived enhancer characteristics found in animals. Our results also point to a strong similarity between coding and non-coding transcription in *Arabidopsis*, as they both appear to initiate and elongate exclusively in one direction.

While the literature describes unique histone mark patterns that distinguish enhancers from promoters, even in their various states of activity, a growing number of studies have also revealed that the line distinguishing these categories has become ever-more blurred (Ernst, Kheradpour et al. 2011, Kowalczyk, Hughes et al. 2012, Kim and Shiekhattar 2015). Increasingly, it seems that the promoter/enhancer divide is a false dichotomy, and future efforts may be better served by searching for unifying characteristics that span *cis*-regulatory elements, rather than minor characteristics that divide them.

The door remains open for other histone PTMs to emerge as the characteristic enhancer signature in *Arabidopsis* with further testing. In a recent cross-species comparison of angiosperms, none of the putative regulatory regions identified matched the characteristic animal enhancer histone PTM profile, and many were devoid of any of the marks assayed for (Lu, Marand et al. 2019). While future investigations may explore other modifications, this study shows that chromatin accessibility data, especially when combined with GRO-seq data, make a compelling case for identifying regulatory regions of interest across the genome. While individual functional testing is needed in the future to further cement these regions' identities as bona fide enhancers, this research opens the door on new, fruitful approaches that aim to accelerate the characterization of the Plantae transcriptional regulatome to the level of its peers.

ACKNOWLEDGEMENTS

We would like to thank the members of the Deal Lab for their feedback and support. We would also like to extend a special thanks to Benjamin Barwick, Ph.D. and Marko Bajic, Ph.D. for their advice regarding sequencing analysis.

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 **Table 4.1**.

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 **Table 4.2**.

Data analysis

Raw sequence read processing, mapping, peak calling, and genomic distribution determination were all conducted as described previously (Maher, Bajic et al. 2018). **Table 4.3** details the data quality of the *A. thaliana* non-hair cell ChIP-seq generated by this study. After processing and mapping, GRO-seq data were run through an R script which removed the top 10% of reads in order to prevent high-signal artifacts from skewing the distribution of the metaplots. Heatmaps and metaplots were generated using SeqPlots (http://seqplots.ga/) (Stempor and Ahringer 2016).

TABLES

Table 4.1: 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	3-day old seedlings	ChIP-seq, ABF3	GEO	GSE80568	GSM2130976	.fastq	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

Table 4.2: *A. thaliana* ChIP-seq antibody information

Target	Antibody Name	Supplier	Concentration	Quantity Used per Reaction
H3K4me1	ab8895	Abcam	0.5 mg/mL	2 μg/4 μL
H3K4me3	ab8580	Abcam	0.45 mg/mL	1.8 μg/4 μL
H3K27ac	ab4729	Abcam	0.5 mg/mL	2 μg/4 μL
H3K27me3	07-449	Millipore	0.5 mg/mL	2 μg/4 μL
H3	ab1791	Abcam	0.5 mg/mL	2 μg/4 μL

Table 4.3: 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

FIGURES

Figure 4.1

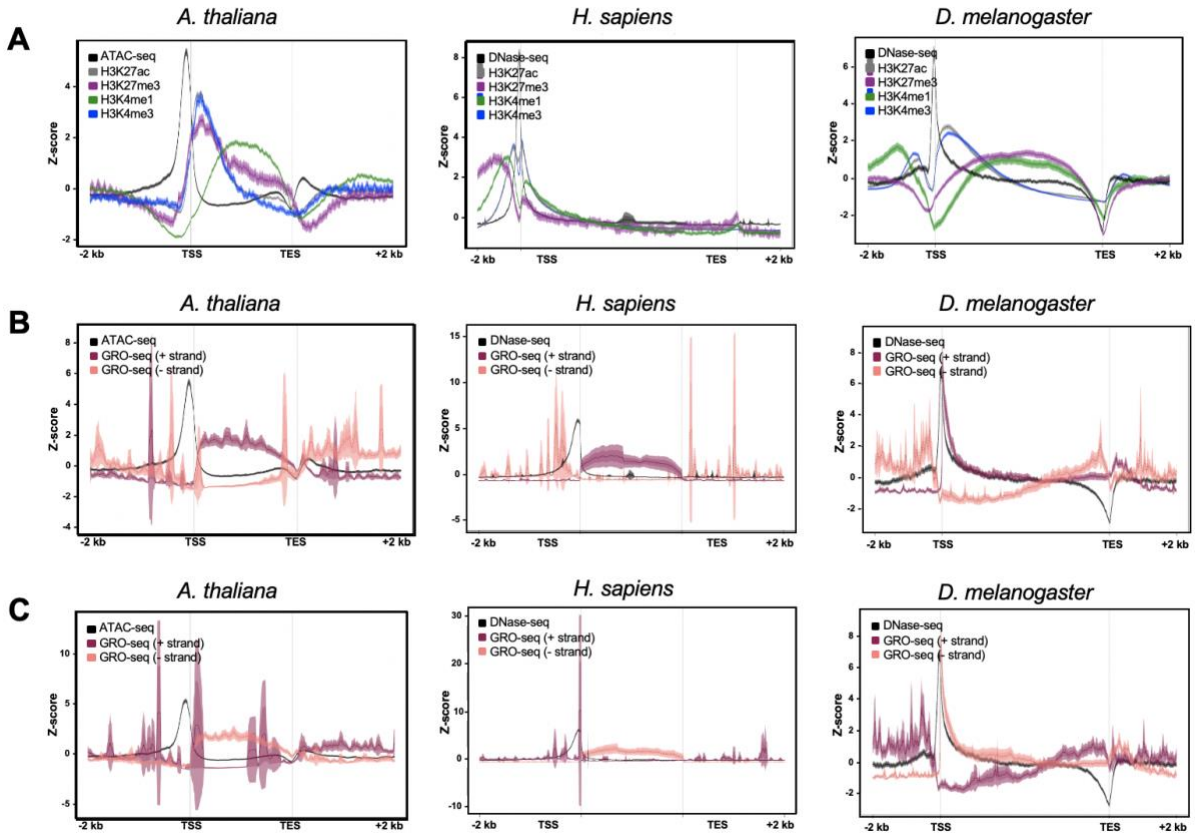


Figure 4.1: Histone modification enrichment and chromatin accessibility across gene bodies reflect global transcriptional distinctions between plants and animals. **A)** Metaplots of average gene profiles of annotated *A. thaliana*, *H. sapiens*, and *D. melanogaster* genes. ChIP-seq signal for H3K27ac, H3K27me3, H3K4me1, and H3K4me3 are shown, as well as chromatin accessibility data (ATAC-seq or DNase-seq). **B)** Metaplots of the nascent transcriptional output (GRO-seq data) on *H. sapiens*, *D. melanogaster*, and *A. thaliana* annotated positive strand and **C)** negative strand genes. Windows extend 2 kb upstream of the transcription start site (TSS) and 2 kb downstream of the transcription end site (TES). For each dataset, several statistical metrics are shown. Solid lines represent the mean signal intensity (in normalized RPKM); the inner, dark-shaded region represents the standard error; and the outer, light-shaded region represents the 95% confidence interval (CI) of signal intensity.

Figure 4.2

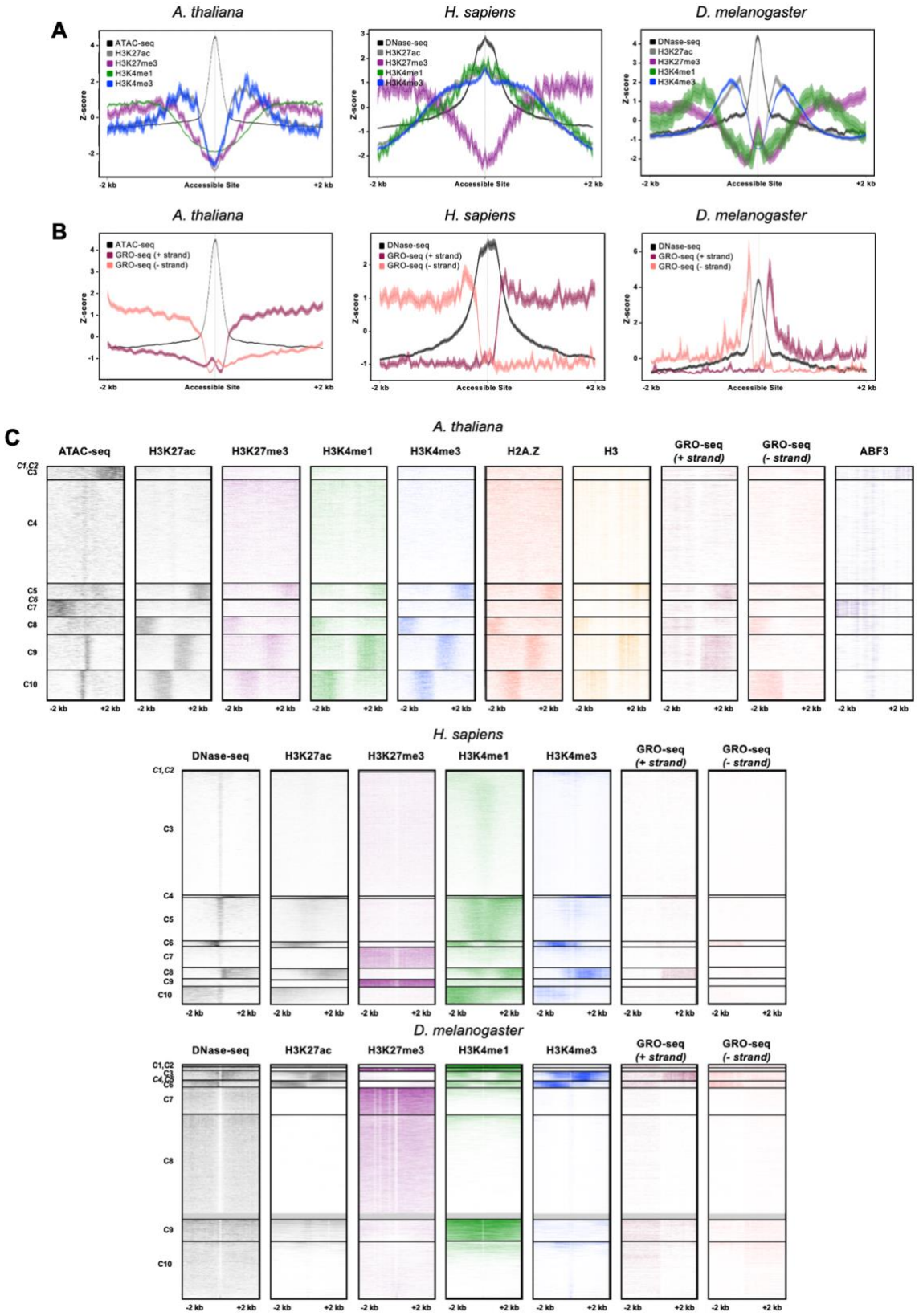

A
A. thaliana
H. sapiens
D. melanogaster
ATAC-seq
H3K27ac
H3K27me3
H3K4me1
H3K4me3
DNase-seq
Z-score
-2 kb
Accessible Site
+2 kb
B
GRO-seq (+ strand)
GRO-seq (- strand)
C
ATAC-seq
H3K27ac
H3K27me3
H3K4me1
H3K4me3
H2A.Z
H3
GRO-seq (+ strand)
GRO-seq (- strand)
ABF3
DNase-seq
C1,C2
C3
C4
C5
C6
C7
C8
C9
C10

Figure 4.2: Enrichment patterns across proximal intergenic accessible chromatin regions. A) Metaplots of average histone modification and chromatin accessibility profiles at proximal intergenic accessible regions in *A. thaliana*, *H. sapiens*, and *D. melanogaster*. **B)** Metaplot of chromatin accessibility and nascent transcriptional output (GRO-seq) at profiles at proximal intergenic accessible regions in *A. thaliana*, *H. sapiens*, and *D. melanogaster*. Solid lines represent the mean signal intensity (in normalized RPKM); 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 *A. thaliana* (27,503 sites), *H. sapiens* (11,778 sites), and *D. melanogaster* (10,038 sites). Heatmaps are divided into 10 k-means clusters; clusters labeled in italics may not be accurately visualized at this resolution. Windows extend 2 kb upstream and 2 kb downstream of the accessible site, and color scale ranges from a z-score of 0-2. Proximal intergenic accessible chromatin regions fall within 2 kb–100 bp upstream of a TSS or 100 bp–1 kb downstream of a TES.

Figure 4.3

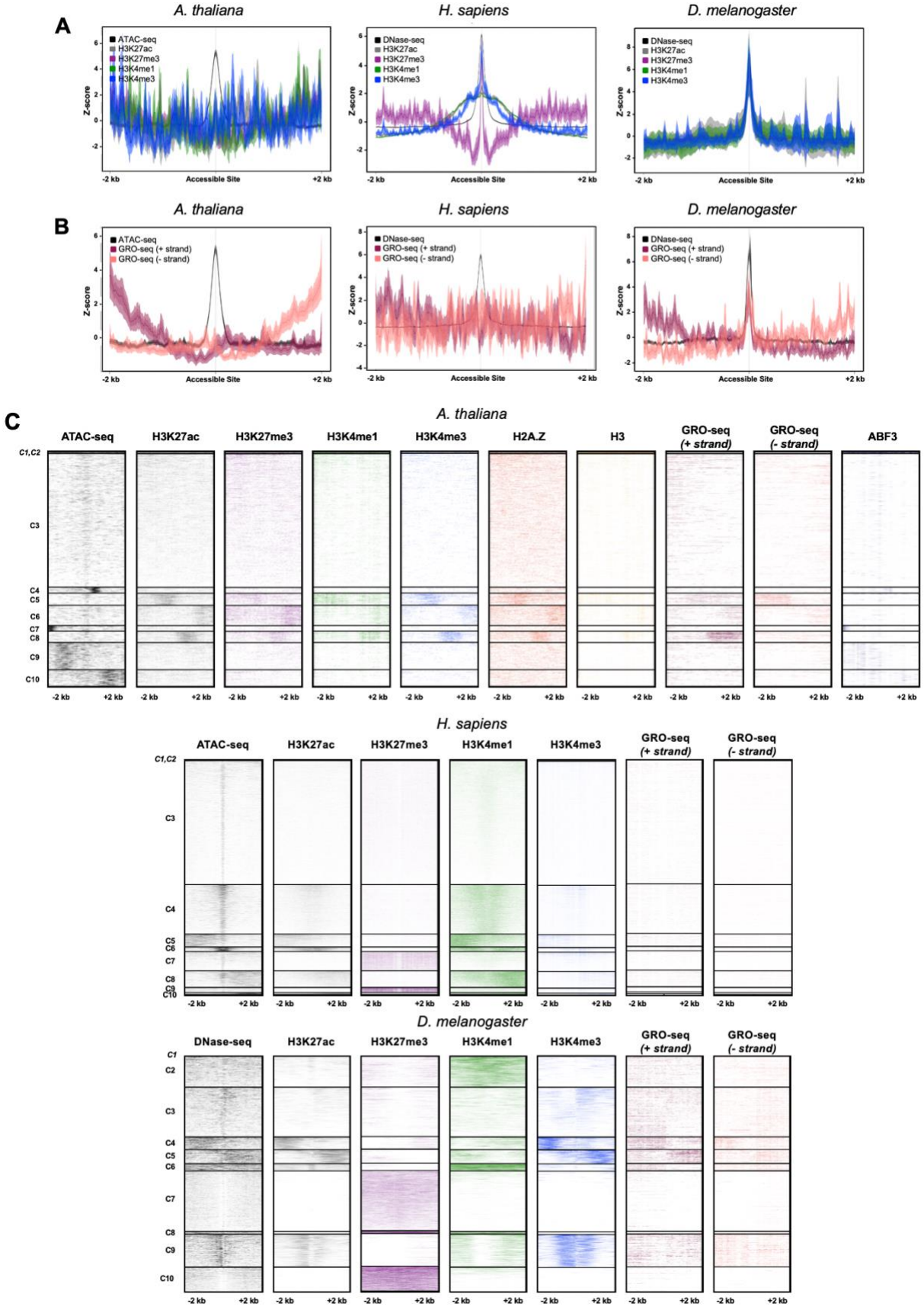
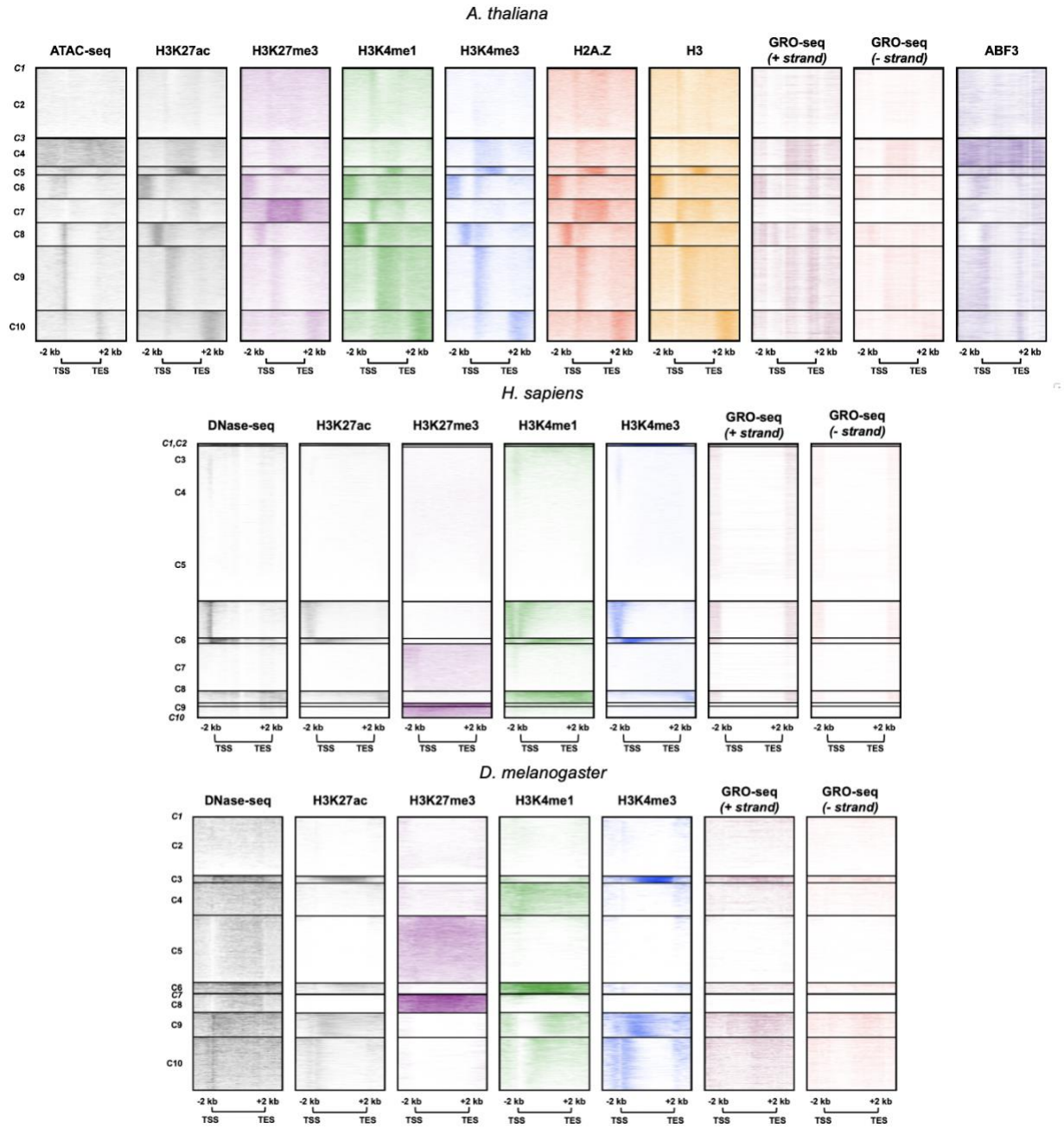

A
A. thaliana
H. sapiens
D. melanogaster
ATAC-seq
H3K27ac
H3K27me3
H3K4me1
H3K4me3
DNase-seq
Z-score
-2 kb
Accessible Site
+2 kb
B
GRO-seq (+ strand)
GRO-seq (- strand)
C
H2A.Z
H3
GRO-seq (+ strand)
GRO-seq (- strand)
ABF3
C1,C2
C3
C4
C5
C6
C7
C8
C9
C10
C1

Figure 4.3: Enrichment patterns across distal intergenic accessible chromatin regions. A) Metaplots of average histone modification and chromatin accessibility profiles at distal intergenic accessible regions in *A. thaliana*, *H. sapiens*, and *D. melanogaster*. **B)** Metaplot of chromatin accessibility and nascent transcriptional output (GRO-seq) at profiles at distal intergenic accessible regions in *A. thaliana*, *H. sapiens*, and *D. melanogaster*. Solid lines represent the mean signal intensity (in normalized RPKM); 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 distal intergenic accessible regions in *A. thaliana* (2,867 sites), *H. sapiens* (11,251 sites), and *D. melanogaster* (5,574 sites). Heatmaps are divided into 10 k-means clusters; clusters labeled in italics may not be accurately visualized at this resolution. Windows extend 2 kb upstream and 2 kb downstream of the accessible site, and color scale ranges from a z-score of 0-2. Distal intergenic accessible chromatin regions fall beyond 2 kb upstream of a TSS and 1 kb downstream of a TES.

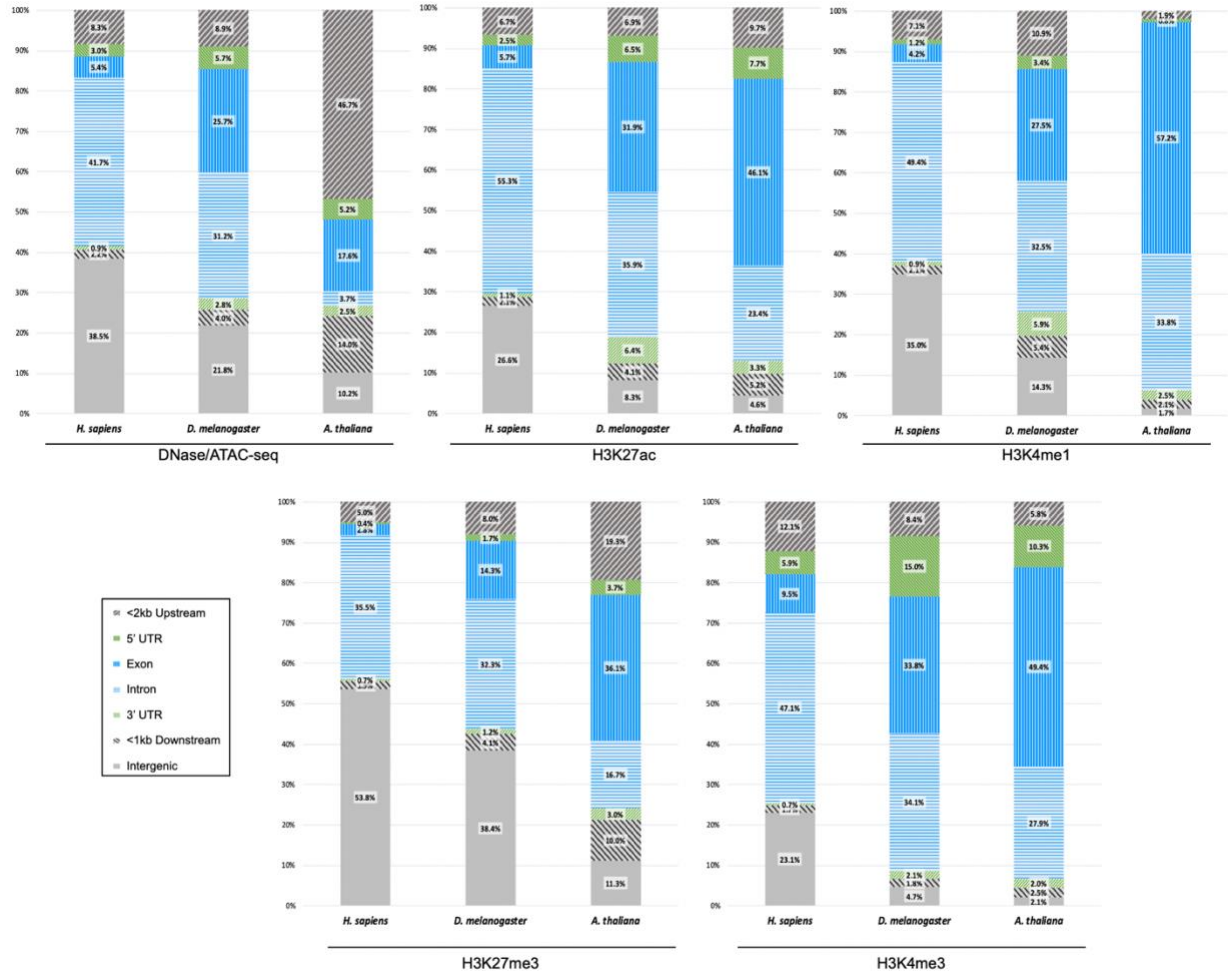
Supplementary Figure 4.1



Supplementary Figure 4.1: Histone modification enrichment and chromatin accessibility across gene bodies reflect transcriptional distinctions between plants and animals in discrete subpopulations. Heatmaps of average gene profiles of annotated *A. thaliana*, *H. sapiens*, and *D. melanogaster* genes. Heatmaps are divided into 10 k-means clusters; clusters labeled in italics may not be accurately visualized

at this resolution. Windows extend 2 kb upstream of the transcription start site (TSS) and 2 kb downstream of the transcription end site (TES), and color scale ranges from a z-score of 0-2.

Supplementary Figure 4.2



Supplementary Figure 4.2: Distribution of histone modification and chromatin accessibility peaks across genomic features. Bar graphs of the distribution of histone PTMs and chromatin accessibility peaks across features of the *H. sapiens*, *D. melanogaster*, and *A. thaliana* genomes.

LITERATURE CITED

(2004). "The ENCODE (ENCyclopedia Of DNA Elements) Project." Science **306**(5696): 636-640.

(2012). "An integrated encyclopedia of DNA elements in the human genome." Nature **489**(7414): 57-74.

Anders, S. and W. Huber (2010). "Differential expression analysis for sequence count data." Genome Biol **11**(10): R106.

Arnold, C. D., D. Gerlach, D. Spies, J. A. Matts, Y. A. Sytnikova, M. Pagani, N. C. Lau and A. Stark (2014). "Quantitative genome-wide enhancer activity maps for five Drosophila species show functional enhancer conservation and turnover during cis-regulatory evolution." Nat Genet **46**(7): 685-692.

Arnold, C. D., D. Gerlach, C. Stelzer, L. M. Boryn, M. Rath and A. Stark (2013). "Genome-wide quantitative enhancer activity maps identified by STARR-seq." Science **339**(6123): 1074-1077.

Barski, A., S. Cuddapah, K. Cui, T. Y. Roh, D. E. Schones, Z. Wang, G. Wei, I. Chepelev and K. Zhao (2007). "High-resolution profiling of histone methylations in the human genome." Cell **129**(4): 823-837.

Bell, O., V. K. Tiwari, N. H. Thoma and D. Schubeler (2011). "Determinants and dynamics of genome accessibility." Nat Rev Genet **12**(8): 554-564.

Berg, P. E., Z. Popovic and W. F. Anderson (1984). "Promoter dependence of enhancer activity." Mol Cell Biol **4**(8): 1664-1668.

Bonn, S., R. P. Zinzen, C. Girardot, E. H. Gustafson, A. Perez-Gonzalez, N. Delhomme, Y. Ghavi-Helm, B. Wilczynski, A. Riddell and E. E. Furlong (2012). "Tissue-specific analysis of chromatin state identifies temporal signatures of enhancer activity during embryonic development." Nat Genet **44**(2): 148-156.

Boyle, A. P., S. Davis, H. P. Shulha, P. Meltzer, E. H. Margulies, Z. Weng, T. S. Furey and G. E. Crawford (2008). "High-resolution mapping and characterization of open chromatin across the genome." Cell **132**(2): 311-322.

Buenrostro, J. D., P. G. Giresi, L. C. Zaba, H. Y. Chang and W. J. Greenleaf (2013). "Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position." Nat Methods **10**(12): 1213-1218.

Buenrostro, J. D., B. Wu, H. Y. Chang and W. J. Greenleaf (2015). "ATAC-seq: A Method for Assaying Chromatin Accessibility Genome-Wide." Curr Protoc Mol Biol **109**: 21.29.21-29.

Chatterjee, S. and N. Ahituv (2017). "Gene Regulatory Elements, Major Drivers of Human Disease." Annu Rev Genomics Hum Genet **18**: 45-63.

Core, L. J., J. J. Waterfall and J. T. Lis (2008). "Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters." Science **322**(5909): 1845-1848.

Creyghton, M. P., A. W. Cheng, G. G. Welstead, T. Kooistra, B. W. Carey, E. J. Steine, J. Hanna, M. A. Lodato, G. M. Frampton, P. A. Sharp, L. A. Boyer, R. A. Young and R. Jaenisch (2010). "Histone H3K27ac separates active from poised enhancers and predicts developmental state." Proc Natl Acad Sci U S A **107**(50): 21931-21936.

Cubenas-Potts, C., M. J. Rowley, X. Lyu, G. Li, E. P. Lei and V. G. Corces (2017). "Different enhancer classes in Drosophila bind distinct architectural proteins and mediate unique chromatin interactions and 3D architecture." Nucleic Acids Res **45**(4): 1714-1730.

Edgar, R., M. Domrachev and A. E. Lash (2002). "Gene Expression Omnibus: NCBI gene expression and hybridization array data repository." Nucleic Acids Res **30**(1): 207-210.

Ernst, J. and M. Kellis (2010). "Discovery and characterization of chromatin states for systematic annotation of the human genome." Nat Biotechnol **28**(8): 817-825.

Ernst, J., P. Kheradpour, T. S. Mikkelsen, N. Shoresh, L. D. Ward, C. B. Epstein, X. Zhang, L. Wang, R. Issner, M. Coyne, M. Ku, T. Durham, M. Kellis and B. E. Bernstein (2011). "Mapping and analysis of chromatin state dynamics in nine human cell types." Nature **473**(7345): 43-49.

Feng, S., S. J. Cokus, V. Schubert, J. Zhai, M. Pellegrini and S. E. Jacobsen (2014). "Genome-wide Hi-C analyses in wild-type and mutants reveal high-resolution chromatin interactions in Arabidopsis." Mol Cell **55**(5): 694-707.

Gardini, A. (2017). "Global Run-On Sequencing (GRO-Seq)." Methods Mol Biol **1468**: 111-120.

Ghavi-Helm, Y., F. A. Klein, T. Pakozdi, L. Ciglar, D. Noordermeer, W. Huber and E. E. Furlong (2014). "Enhancer loops appear stable during development and are associated with paused polymerase." Nature **512**(7512): 96-100.

Grossniklaus, U. and R. Paro (2014). "Transcriptional silencing by polycomb-group proteins." Cold Spring Harb Perspect Biol **6**(11): a019331.

Guenther, M. G., S. S. Levine, L. A. Boyer, R. Jaenisch and R. A. Young (2007). "A chromatin landmark and transcription initiation at most promoters in human cells." Cell **130**(1): 77-88.

Hah, N., S. Murakami, A. Nagari, C. G. Danko and W. L. Kraus (2013). "Enhancer transcripts mark active estrogen receptor binding sites." Genome Res **23**(8): 1210-1223.

Hawkins, R. D., G. C. Hon, L. K. Lee, Q. Ngo, R. Lister, M. Pelizzola, L. E. Edsall, S. Kuan, Y. Luu, S. Klugman, J. Antosiewicz-Bourget, Z. Ye, C. Espinoza, S. Agarwahl, L. Shen, V. Ruotti, W. Wang, R. Stewart, J. A. Thomson, J. R. Ecker and B. Ren (2010). "Distinct epigenomic landscapes of pluripotent and lineage-committed human cells." Cell Stem Cell **6**(5): 479-491.

Heintzman, N. D., G. C. Hon, R. D. Hawkins, P. Kheradpour, A. Stark, L. F. Harp, Z. Ye, L. K. Lee, R. K. Stuart, C. W. Ching, K. A. Ching, J. E. Antosiewicz-Bourget, H. Liu, X. Zhang, R. D. Green, V. V. Lobanenkov, R. Stewart, J. A. Thomson, G. E. Crawford, M. Kellis and B. Ren (2009). "Histone modifications at human enhancers reflect global cell-type-specific gene expression." Nature **459**(7243): 108-112.

Heintzman, N. D., R. K. Stuart, G. Hon, Y. Fu, C. W. Ching, R. D. Hawkins, L. O. Barrera, S. Van Calcar, C. Qu, K. A. Ching, W. Wang, Z. Weng, R. D. Green, G. E. Crawford and B. Ren (2007). "Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome." Nat Genet **39**(3): 311-318.

Henikoff, S., J. G. Henikoff, A. Sakai, G. B. Loeb and K. Ahmad (2009). "Genome-wide profiling of salt fractions maps physical properties of chromatin." Genome Res **19**(3): 460-469.

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

Jin, C., C. Zang, G. Wei, K. Cui, W. Peng, K. Zhao and G. Felsenfeld (2009). "H3.3/H2A.Z double variant-containing nucleosomes mark 'nucleosome-free regions' of active promoters and other regulatory regions." Nat Genet **41**(8): 941-945.

Kaikkonen, M. U., N. J. Spann, S. Heinz, C. E. Romanoski, K. A. Allison, J. D. Stender, H. B. Chun, D. F. Tough, R. K. Prinjha, C. Benner and C. K. Glass (2013). "Remodeling of the enhancer landscape during macrophage activation is coupled to enhancer transcription." Mol Cell **51**(3): 310-325.

Kapranov, P., J. Cheng, S. Dike, D. A. Nix, R. Duttagupta, A. T. Willingham, P. F. Stadler, J. Hertel, J. Hackermuller, I. L. Hofacker, I. Bell, E. Cheung, J. Drenkow, E. Dumais, S. Patel, G. Helt, M. Ganesh, S. Ghosh, A. Piccolboni, V. Sementchenko, H. Tammana and T. R. Gingeras (2007). "RNA maps reveal new RNA classes and a possible function for pervasive transcription." Science **316**(5830): 1484-1488.

Keene, M. A., V. Corces, K. Lowenhaupt and S. C. Elgin (1981). "DNase I hypersensitive sites in Drosophila chromatin occur at the 5' ends of regions of transcription." Proc Natl Acad Sci U S A **78**(1): 143-146.

Kim, T. H., L. O. Barrera, M. Zheng, C. Qu, M. A. Singer, T. A. Richmond, Y. Wu, R. D. Green and B. Ren (2005). "A high-resolution map of active promoters in the human genome." Nature **436**(7052): 876-880.

Kim, T. K., M. Hemberg, J. M. Gray, A. M. Costa, D. M. Bear, J. Wu, D. A. Harmin, M. Laptewicz, K. Barbara-Haley, S. Kuersten, E. Markenscoff-Papadimitriou, D. Kuhl, H. Bito, P. F. Worley, G. Kreiman and M. E. Greenberg (2010). "Widespread transcription at neuronal activity-regulated enhancers." Nature **465**(7295): 182-187.

Kim, T. K. and R. Shiekhattar (2015). "Architectural and Functional Commonalities between Enhancers and Promoters." Cell **162**(5): 948-959.

Kleinjan, D. A. and V. van Heyningen (2005). "Long-range control of gene expression: emerging mechanisms and disruption in disease." Am J Hum Genet **76**(1): 8-32.

Kowalczyk, M. S., J. R. Hughes, D. Garrick, M. D. Lynch, J. A. Sharpe, J. A. Sloane-Stanley, S. J. McGowan, M. De Gobbi, M. Hosseini, D. Vernimmen, J. M. Brown, N. E. Gray, L. Collavin, R. J. Gibbons, J. Flint, S. Taylor, V. J. Buckle, T. A. Milne, W. G. Wood and D. R. Higgs (2012). "Intragenic enhancers act as alternative promoters." Mol Cell **45**(4): 447-458.

Kvon, E. Z., T. Kazmar, G. Stampfel, J. O. Yanez-Cuna, M. Pagani, K. Schernhuber, B. J. Dickson and A. Stark (2014). "Genome-scale functional characterization of Drosophila developmental enhancers in vivo." Nature **512**(7512): 91-95.

Lai, F. and R. Shiekhattar (2014). "Enhancer RNAs: the new molecules of transcription." Curr Opin Genet Dev **25**: 38-42.

Lee, L. R., D. L. Wengier and D. C. Bergmann (2019). "Cell-type-specific transcriptome and histone modification dynamics during cellular reprogramming in the Arabidopsis stomatal lineage." Proc Natl Acad Sci U S A **116**(43): 21914-21924.

Lee, T. I. and R. A. Young (2000). "Transcription of eukaryotic protein-coding genes." Annu Rev Genet **34**: 77-137.

Lettice, L. A., S. J. Heaney, L. A. Purdie, L. Li, P. de Beer, B. A. Oostra, D. Goode, G. Elgar, R. E. Hill and E. de Graaff (2003). "A long-range Shh enhancer regulates expression in the developing limb and fin and is associated with preaxial polydactyly." Hum Mol Genet **12**(14): 1725-1735.

Li, B., M. Carey and J. L. Workman (2007). "The role of chromatin during transcription." Cell **128**(4): 707-719.

Liu, J., C. Jung, J. Xu, H. Wang, S. Deng, L. Bernad, C. Arenas-Huertero and N. H. Chua (2012). "Genome-wide analysis uncovers regulation of long intergenic noncoding RNAs in Arabidopsis." Plant Cell **24**(11): 4333-4345.

Lu, Z., A. P. Marand, W. A. Ricci, C. L. Ethridge, X. Zhang and R. J. Schmitz (2019). "The prevalence, evolution and chromatin signatures of plant regulatory elements." Nat Plants **5**(12): 1250-1259.

Maher, K. A., M. Bajic, K. Kajala, M. Reynoso, G. Pauluzzi, D. A. West, K. Zumstein, M. Woodhouse, K. Bubb, M. W. Dorrity, C. Queitsch, J. Bailey-Serres, N. Sinha, S. M. Brady and R. B. Deal (2018).

"Profiling of Accessible Chromatin Regions across Multiple Plant Species and Cell Types Reveals Common Gene Regulatory Principles and New Control Modules." Plant Cell **30**(1): 15-36.

McGhee, J. D., W. I. Wood, M. Dolan, J. D. Engel and G. Felsenfeld (1981). "A 200 base pair region at the 5' end of the chicken adult beta-globin gene is accessible to nuclease digestion." Cell **27**(1 Pt 2): 45-55.

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

Mendenhall, E. M., K. E. Williamson, D. Reyon, J. Y. Zou, O. Ram, J. K. Joung and B. E. Bernstein (2013). "Locus-specific editing of histone modifications at endogenous enhancers." Nat Biotechnol **31**(12): 1133-1136.

Muerdter, F., L. M. Boryn and C. D. Arnold (2015). "STARR-seq - principles and applications." Genomics **106**(3): 145-150.

Mumbach, M. R., A. T. Satpathy, E. A. Boyle, C. Dai, B. G. Gowen, S. W. Cho, M. L. Nguyen, A. J. Rubin, J. M. Granja, K. R. Kazane, Y. Wei, T. Nguyen, P. G. Greenside, M. R. Corces, J. Tycko, D. R. Simeonov, N. Suliman, R. Li, J. Xu, R. A. Flynn, A. Kundaje, P. A. Khavari, A. Marson, J. E. Corn, T. Quertermous, W. J. Greenleaf and H. Y. Chang (2017). "Enhancer connectome in primary human cells identifies target genes of disease-associated DNA elements." Nat Genet **49**(11): 1602-1612.

Oka, R., J. Zicola, B. Weber, S. N. Anderson, C. Hodgman, J. I. Gent, J. J. Wesselink, N. M. Springer, H. C. J. Hoefsloot, F. Turck and M. Stam (2017). "Genome-wide mapping of transcriptional enhancer candidates using DNA and chromatin features in maize." Genome Biol **18**(1): 137.

Ong, C. T. and V. G. Corces (2011). "Enhancer function: new insights into the regulation of tissue-specific gene expression." Nat Rev Genet **12**(4): 283-293.

Pajoro, A., P. Madrigal, J. M. Muino, J. T. Matus, J. Jin, M. A. Mecchia, J. M. Debernardi, J. F. Palatnik, S. Balazadeh, M. Arif, D. S. O'Maoileidigh, F. Wellmer, P. Krajewski, J. L. Riechmann, G. C. Angenent and K. Kaufmann (2014). "Dynamics of chromatin accessibility and gene regulation by MADS-domain transcription factors in flower development." Genome Biol **15**(3): R41.

Rada-Iglesias, A., R. Bajpai, T. Swigut, S. A. Brugmann, R. A. Flynn and J. Wysocka (2011). "A unique chromatin signature uncovers early developmental enhancers in humans." Nature **470**(7333): 279-283.

Schones, D. E., K. Cui, S. Cuddapah, T. Y. Roh, A. Barski, Z. Wang, G. Wei and K. Zhao (2008). "Dynamic regulation of nucleosome positioning in the human genome." Cell **132**(5): 887-898.

Schwaiger, M., A. Schonauer, A. F. Rendeiro, C. Pribitzer, A. Schauer, A. F. Gilles, J. B. Schinko, E. Renfer, D. Fredman and U. Technau (2014). "Evolutionary conservation of the eumetazoan gene regulatory landscape." Genome Res **24**(4): 639-650.

Seila, A. C., J. M. Calabrese, S. S. Levine, G. W. Yeo, P. B. Rahl, R. A. Flynn, R. A. Young and P. A. Sharp (2008). "Divergent transcription from active promoters." Science **322**(5909): 1849-1851.

Seila, A. C., L. J. Core, J. T. Lis and P. A. Sharp (2009). "Divergent transcription: a new feature of active promoters." Cell Cycle **8**(16): 2557-2564.

Shilatifard, A. (2006). "Chromatin modifications by methylation and ubiquitination: implications in the regulation of gene expression." Annu Rev Biochem **75**: 243-269.

Shlyueva, D., G. Stampfel and A. Stark (2014). "Transcriptional enhancers: from properties to genome-wide predictions." Nat Rev Genet **15**(4): 272-286.

Spitz, F. and E. E. Furlong (2012). "Transcription factors: from enhancer binding to developmental control." Nat Rev Genet **13**(9): 613-626.

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

Torres, E. S. and R. B. Deal (2019). "The histone variant H2A.Z and chromatin remodeler BRAHMA act coordinately and antagonistically to regulate transcription and nucleosome dynamics in Arabidopsis." Plant J **99**(1): 144-162.

Trinklein, N. D., S. F. Aldred, S. J. Hartman, D. I. Schroeder, R. P. Otillar and R. M. Myers (2004). "An abundance of bidirectional promoters in the human genome." Genome Res **14**(1): 62-66.

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

Vernimmen, D. and W. A. Bickmore (2015). "The Hierarchy of Transcriptional Activation: From Enhancer to Promoter." Trends Genet **31**(12): 696-708.

Villar, D., C. Berthelot, S. Aldridge, T. F. Rayner, M. Lukk, M. Pignatelli, T. J. Park, R. Deaville, J. T. Erichsen, A. J. Jasinska, J. M. Turner, M. F. Bertelsen, E. P. Murchison, P. Flicek and D. T. Odom (2015). "Enhancer evolution across 20 mammalian species." Cell **160**(3): 554-566.

Wang, C., C. Liu, D. Roqueiro, D. Grimm, R. Schwab, C. Becker, C. Lanz and D. Weigel (2015). "Genome-wide analysis of local chromatin packing in Arabidopsis thaliana." Genome Res **25**(2): 246-256.

Wang, D. and R. B. Deal (2015). "Epigenome profiling of specific plant cell types using a streamlined INTACT protocol and ChIP-seq." Methods Mol Biol **1284**: 3-25.

Wang, H., P. J. Chung, J. Liu, I. C. Jang, M. J. Kean, J. Xu and N. H. Chua (2014). "Genome-wide identification of long noncoding natural antisense transcripts and their responses to light in Arabidopsis." Genome Res **24**(3): 444-453.

Wang, M., L. Tu, M. Lin, Z. Lin, P. Wang, Q. Yang, Z. Ye, C. Shen, J. Li, L. Zhang, X. Zhou, X. Nie, Z. Li, K. Guo, Y. Ma, C. Huang, S. Jin, L. Zhu, X. Yang, L. Min, D. Yuan, Q. Zhang, K. Lindsey and X. Zhang (2017). "Asymmetric subgenome selection and cis-regulatory divergence during cotton domestication." Nat Genet **49**(4): 579-587.

Wang, Z., C. Zang, J. A. Rosenfeld, D. E. Schones, A. Barski, S. Cuddapah, K. Cui, T. Y. Roh, W. Peng, M. Q. Zhang and K. Zhao (2008). "Combinatorial patterns of histone acetylations and methylations in the human genome." Nat Genet **40**(7): 897-903.

Weber, B., J. Zicola, R. Oka and M. Stam (2016). "Plant Enhancers: A Call for Discovery." Trends Plant Sci **21**(11): 974-987.

Xu, H. and T. R. Hoover (2001). "Transcriptional regulation at a distance in bacteria." Curr Opin Microbiol **4**(2): 138-144.

Yan, W., D. Chen, J. Schumacher, D. Durantini, J. Engelhorn, M. Chen, C. C. Carles and K. Kaufmann (2019). "Dynamic control of enhancer activity drives stage-specific gene expression during flower morphogenesis." Nat Commun **10**(1): 1705.

Zentner, G. E., P. J. Tesar and P. C. Scacheri (2011). "Epigenetic signatures distinguish multiple classes of enhancers with distinct cellular functions." Genome Res **21**(8): 1273-1283.

Zhang, W., Y. Wu, J. C. Schnable, Z. Zeng, M. Freeling, G. E. Crawford and J. Jiang (2012). "High-resolution mapping of open chromatin in the rice genome." Genome Res **22**(1): 151-162.

Zhang, W., T. Zhang, Y. Wu and J. Jiang (2012). "Genome-wide identification of regulatory DNA elements and protein-binding footprints using signatures of open chromatin in Arabidopsis." Plant Cell **24**(7): 2719-2731.

Zhu, B., W. Zhang and T. Zhang (2015). "Genome-Wide Prediction and Validation of Intergenic Enhancers in Arabidopsis Using Open Chromatin Signatures." **27**(9): 2415-2426.

CHAPTER 5: DISCUSSION – CONCLUSIONS AND FUTURE DIRECTIONS

In this dissertation, I have shown a variety of approaches designed to identify and examine the qualities of enhancer elements in plant species. First, by the union of ATAC-seq with nuclei-purifying techniques, including sucrose sedimentation and INTACT, I helped design a methodology for generating rapid data on chromatin accessibility with little to no interfering signal from organellar DNA. Through the use of this approach we investigated the broad conservation of the transcriptional regulatory landscape across species, as well as the finer differences in open chromatin landscapes between developmentally related root cell types. In our cross-species comparison between *Arabidopsis*, *Medicago*, tomato, and rice, we found that the majority of transposase hypersensitive sites (THSs) are found within the 3 kb upstream of the transcription start site (TSS), regardless of the size of the genome or the proportion of it taken up by intergenic space. Despite a lack of similar open chromatin profiles between orthologs and expressologs (i.e. functional homologs), the transcription factors that regulate these gene sets appear to be evolutionarily conserved, though the order and location of their binding motifs is not. The comparison of the open chromatin profile of hair and non-hair root cells revealed global similarities, indicating that THSs alone – even those differentially accessible between cell types – cannot be used to reliably distinguish between transcriptionally activating and transcriptionally repressive binding events. Nonetheless, transcription factor motif analysis illuminated the basis for a new hair cell fate developmental control module driven by a MYB signaling feedback loop which merits further investigation. Finally, our comparison of histone modifications and transcriptional output at putative regulatory regions across plant and animal genomes revealed an overwhelming bias for unidirectional transcription in *Arabidopsis*, suggesting that the process of RNA polymerase II initiation may be mechanistically less promiscuous than in humans and flies.

Our work raises a number of critical questions on the nature of transcriptional regulation and its component parts. In this section I further explore our cumulative findings, how they relate to the latest discoveries in the field, and possibilities for future approaches to enhancer discovery across the plant kingdom.

The Distinction between Enhancers and Other Cis-*Regulatory Elements is Ambiguous*

The first description of enhancers emerged in the early 1980s, originating with histone H2A gene transcription in sea urchins (Grosschedl and Birnstiel 1980), significantly expanding with the characterization of the viral SV40 enhancer (Banerji, Rusconi et al. 1981, Benoist and Chambon 1981, Moreau, Hen et al. 1981, Fromm and Berg 1982, Fromm and Berg 1983, Khoury and Gruss 1983, Atchison 1988) and the mammalian immunoglobulin heavy-chain enhancer (Banerji, Olson et al. 1983, Gillies, Morrison et al. 1983, Neuberger 1983). These foundational studies defined the criteria that are still used to describe this regulatory element class. Namely, enhancers are sequences of DNA which amplify the transcriptional output of their target gene(s) that can function in an orientation- and distance-independent manner (Kim and Shiekhattar 2015).

These traits are often used to distinguish enhancer candidates from a similar regulatory element class: genetic promoters. Promoters are defined as sequences that can autonomously initiate transcription (Mikhaylichenko, Bondarenko et al. 2018). Core promoters are positioned proximally upstream of their target gene, within the ~100 bp surrounding the transcription start site (TSS). They can be identified by characteristic sequence motifs, including the TATA box (Lifton, Goldberg et al. 1978, FitzGerald, Sturgill et al. 2006), CpG islands (CGIs) (Gardiner-Garden and Frommer 1987, Saxonov, Berg et al. 2006), Initiator (Inr) motif (Smale and Baltimore 1989), and the downstream promoter elements (DPE) (Burke and Kadonaga 1996). These motifs recruit RNA Polymerase II (RNAPII) and general transcription factors (GTFs), and ultimately act as a binding site for the assembly of the transcriptional pre-initiation complex (PIC) (Hampsey 1998). Promoter elements are thought to only produce stable transcripts in the 5'-3' orientation, and while they are sufficient to initiate transcription of their gene target on their own, they are usually only capable of producing transcripts at a basal level (Kadonaga 2012).

In contrast, enhancers have no similar set of characteristic sequence motifs that are conserved between species. Rather than residing directly upstream, enhancer elements have been found in a variety of locations with regards to their target promoter – upstream; downstream; within introns; and extremely

distally, in some cases megabases away from their target or on entirely separate chromosomes (Lettice, Heaney et al. 2003, Kleinjan and van Heyningen 2005, Ong and Corces 2011). Historically, enhancers have played a supporting role in transcription; thought to be unable to initiate transcription themselves, these elements recruit and bind transcription factors (TFs) in order to amplify the transcriptional output of a promoter-bound PIC.

Together, these criteria outline distinct, non-overlapping, and complementary roles for promoters and enhancers in the process of transcription. As is often the case with biology, however, as these regulatory elements become more closely studied the less clear the distinctions become. As the current state of the field stands, how similar have promoters and enhancers been revealed to be? Do these shifts in dogma cumulate to a relatively minor adjustment, or is a critical rethinking of the categories themselves called for? Below many of the most notable recent discoveries have been summarized.

First, a fundamental and perhaps unsurprising similarity between promoters and enhancers is a shared hypersensitivity to DNase I digestion, an enrichment of H2A.Z/H3.3 nucleosomes within the element, and well-positioned, phased nucleosomes flanking the site (Yuan, Liu et al. 2005, Mavrich, Jiang et al. 2008, Jiang and Pugh 2009, Jin, Zang et al. 2009). This finding fits neatly with the elements' purpose in binding *trans*-acting factors, as nucleosomes comprised of these and other histone variants have been found to be highly dynamic, thus allowing easier access to the underlying DNA (Jin, Zang et al. 2009, Mieczkowski, Cook et al. 2016, Mueller, Mieczkowski et al. 2017).

Beyond their chromatin being accessible, many similarities have been found in the types of factors that bind to these regulatory regions. In addition to TFs and cofactors (Andersson, Gebhard et al. 2014), enhancers act as a binding site for general transcription factors (GTFs) and RNAPII itself, assembling their own PICs (Koch, Fenouil et al. 2011, Bonn, Zinzen et al. 2012, Lai and Pugh 2017). It has been suggested that the assembly of the transcriptional machinery on enhancers is in preparation for it to be transferred onto its target promoter (Moreau, Hen et al. 1981). While the enrichment in the local microenvironment for GTFs may assist in the assembly of the promoter PIC, enhancers themselves have been found to be enriched for both the unphosphorylated, pre-initiation form of RNAPII and the Ser-5-phosphorylated, post-

promoter-escape forms of RNAPII (Natoli and Andrau 2012). This suggests that the holoenzyme is functionally active and is undergoing transcriptional initiation at these regions, a finding that ultimately led to the discovery of a new subspecies of RNA, called 'enhancer RNA' (eRNA).

Enhancer RNAs fall largely into two categories; the majority of eRNAs have been found to be short (< 2 kb), transcribed bidirectionally, and lack polyadenylated tails (Kim, Hemberg et al. 2010, 2012, Djebali, Davis et al. 2012, Harrow, Frankish et al. 2012). However, many instances of long, unidirectional, and polyadenylated eRNAs have been reported (De Santa, Barozzi et al. 2010, Derrien, Johnson et al. 2012). Interestingly, transcriptional bidirectionality and a lack of polyadenylation appear to be correlated – with unidirectionality and polyadenylation being similarly correlated. However, the functional purpose of these distinct characteristics, if any, remains unclear (Koch, Fenouil et al. 2011, Natoli and Andrau 2012). Moreover, whether or not eRNAs have a functional purpose at large remains widely debated. Some suggest that intergenic transcription is stochastic and random, and occurs simply because – from an evolutionary perspective – it would be too energetically costly to suppress the activity of RNAPII at all possible off-target, degenerate promoter sequences across the genome (Struhl 2007). Others point out that the transcriptional activity of RNAPII itself disrupts nucleosomes and keeps chromatin in an accessible state. As such, the production of transcripts at enhancers could be no more than the inconsequential byproduct of a mechanism to maintain accessibility for *trans*-acting factors (Gilchrist, Dos Santos et al. 2010, Mousavi, Zare et al. 2013, Scruggs, Gilchrist et al. 2015). There has been other evidence, however, suggesting that eRNAs have functional activity. Enhancer RNAs may recruit and bind transcription factors to create an activating microenvironment (Muerdter and Stark 2016, Hnisz, Shrinivas et al. 2017, Tsai, Muthusamy et al. 2017), similar to what has been described for P granules in the germline (Brangwynne, Eckmann et al. 2009). Other studies have shown eRNAs facilitating looping between enhancer-promoter pairs by recruiting cohesion (Li, Notani et al. 2013) and the Mediator complex (Hsieh, Fei et al. 2014) to the association site. Furthermore, eRNAs may also contribute to relieving RNAPII pausing by competing for binding with negative elongation factor (NELF), allowing RNAPII to proceed into transcriptional elongation from the promoter (Schaukowitch, Joo et al. 2014).

In this same vein, it has been shown that transcription at promoters is not as tightly controlled as was previously believed. Mammalian promoters show an abundance of bidirectional transcription, assembling two independent RNAPII complexes at the TSS in divergent directions (Core, Waterfall et al. 2008, Seila, Calabrese et al. 2008, Core, Martins et al. 2014). In addition to producing mRNA in the sense direction, promoters produce short, non-coding RNAs in the antisense direction. These molecules are unstable and short-lived, and are degraded through the exosome pathway similar to eRNAs (Preker, Nielsen et al. 2008). These anti-sense transcriptional regions are often enriched in enhancer-like histone modifications, such as H3K4me1, while H3K27ac can be found on nucleosomes both upstream and downstream of the TSS (Barski, Cuddapah et al. 2007, Scruggs, Gilchrist et al. 2015). In addition to producing divergent transcripts like enhancer elements, several cases of promoters functioning as enhancers have come to light (Arnold, Gerlach et al. 2013, Zabidi, Arnold et al. 2015, Nguyen, Jones et al. 2016, Arnold, Zabidi et al. 2017, Dao, Galindo-Albarran et al. 2017). These include promoters being able to act as an inducible enhancer when cloned downstream of a target gene (Serfling, Lubbe et al. 1985), as well as promoters associating in 3D space with other promoters to activate transcription, much as an enhancer would associate with its target promoter (Li, Ruan et al. 2012). Conversely, enhancers have been found to function as promoters (Nguyen, Jones et al. 2016, van Arensbergen, FitzPatrick et al. 2017, Mikhaylichenko, Bondarenko et al. 2018), including intragenic enhancers acting as tissue-specific alternative promoters, producing abundant, spliced, multi-exonic, poly-adenylated RNAs (Kowalczyk, Hughes et al. 2012).

While the functional criteria that distinguish enhancers and promoters appears to widely overlap, other traits have been described that differentiate the element classes. Histone posttranslational modifications (PTMs) are placed on nucleosomes by a variety of ‘writer’ chromatin machinery, and have been used to tease apart a variety of underlying genomic sequences, epigenetic states, and environmental responses. Many histone PTMs have been found to be associated with enhancers in a variety of species, including H3K9me1 (Barski, Cuddapah et al. 2007, Wang, Zang et al. 2008), H3K18ac (Wang, Zang et al. 2008), H3K4me2 (Ernst, Kheradpour et al. 2011), H3K9ac and H3K14ac (Roh, Cuddapah et al. 2005, Roh,

Wei et al. 2007). One set of modifications, however, has been shown to be the most consistently conserved across eukaryotic species. Enhancers are most commonly associated with methylation of lysine 4 and methylation or acetylation marks on lysine 27 of histone 3. 'Active enhancers' bear H3K27ac and H3K4me1, 'poised enhancers' have H3K27me3 and H3K4me1, and 'intermediate enhancers' are associated with H3K4me1 alone (Creyghton, Cheng et al. 2010, Rada-Iglesias, Bajpai et al. 2011). 'Active promoters', on the other hand, are enriched for H3K4me3, with 'poised/bivalent promoters' enriched for both transcriptionally-activating H3K4me3 and the transcriptionally-repressive Polycomb mark H3K27me3 (Bernstein, Mikkelsen et al. 2006).

Overlap in the H3K4me1/H3K4me3 enrichment in both element classes has been described. As such, typically a higher ratio of H3K4me1:H3K4me3 enrichment has been used as a guideline to distinguish enhancers from promoters (Heintzman, Stuart et al. 2007, Heintzman, Hon et al. 2009, Kim and Shiekhattar 2015). Though useful as a general rule of thumb, this guideline seems to be imperfect. Several studies suggest that the mono-, di-, and trimethylation of lysine 4 of histone 3 has more to do with the level of transcription than the type of underlying element itself. Regions with a higher degree of transcriptional output become enriched with H3K4me3 as the transcriptional machinery repeatedly associates with the area, depositing more and more methylation marks to maintain the open chromatin state. Regions that are more infrequently transcribed, such as enhancers, receive comparatively less methyl deposition, and are more likely to be enriched for H3K4me1/2 marks. Following this line of reasoning, it is not surprising that the most active/highly transcribed enhancers have been found to be enriched for H3K4me3, much like promoters (Koch, Fenouil et al. 2011, Pekowska, Benoukraf et al. 2011, Core, Martins et al. 2014). Elongation-specific marks, including Ser-2-phosphorylated RNAPII and H3K36me3 have not been found at actively transcribed enhancer regions (Bonn, Zinzen et al. 2012, Djebali, Davis et al. 2012), though this is likely because these regions are too short to accumulate anything beyond initiation-specific marks (Kim and Shiekhattar 2015).

This degree of overlap between promoter and enhancer characteristics is not entirely surprising. Increasingly, emerging evidence points to promoters and enhancers being ends on the same evolutionary

continuum. Because of their low information content, degenerate forms of TATA-box or Inr motifs are easy to find throughout the genome, and are often enriched around enhancer TSSs and promoter antisense TSSs (Andersson, Gebhard et al. 2014, Core, Martins et al. 2014, Scruggs, Gilchrist et al. 2015). The more closely these degenerate motifs match the consensus sequence, the more promoter-like activity the enhancer elements possess (Mikhaylichenko, Bondarenko et al. 2018). In accordance with the model of fortuitous initiation, bidirectional transcription is the default at new transcriptional regions, and it is only over time that one direction or the other may become repressed or amplified through molecular evolutionary selection (Jin, Eser et al. 2017). This suggests that all promoters were once weak enhancers, all enhancers may be on their way to becoming promoters, and that most regulatory elements lie somewhere within the gray area of this spectrum. Regardless where this particular line of research ultimately leads, it is clear that the working definitions of 'promoter' and 'enhancer', as used by the genomics community, need revisiting.

Histone Modifications in Plant Species

Our recent findings in *Arabidopsis* and other plant species complement these emerging trends in animal studies. When we measured the average distance of open chromatin sites to their nearest transcription start site (TSS) in *A. thaliana*, *M. truncatula*, *O. sativa*, and *S. lycopersicum*, it was found that the majority of the sites fall in the proximal upstream region, < 3 kb away from the TSS (**Chapter 3**). This is in sharp contrast to the distance between enhancer-promoter pairs in *Drosophila* and humans, which averages to tens (Kvon, Kazmar et al. 2014) or hundreds of kilobases away (Mumbach, Satpathy et al. 2017), respectively. As such, *cis*-regulatory elements in plant species are much closer to gene bodies on average than in animal genomes, occupying distances that are far more in line with proximal promoter elements (< 1 kb) than their animal counterparts.

Furthermore, the abundance of hair cell TFs at both highly expressed hair cell genes and highly expressed non-hair cell genes makes the delineation of their function more ambiguous (**Chapter 3**). Rather than strictly categorizing transcription factors as 'activators' or 'repressors', their role may be more context-dependent. This, in turn, suggests that function of the *cis*-regulatory element at large may be context-

dependent instead of inherent in the sequence itself, rendering the terms of 'enhancer' and 'silencer' as operational, rather than categorical. Additionally, just as our chromatin accessibility data revealed global similarities between the profiles of *Arabidopsis* root tip, non-hair, and hair cells, analogous similarities were found in shoot tissue between terminally differentiated mesophyll cells and shoot apical meristem cells (i.e. stem cells) (Sijacic, Bajic et al. 2018). This tendency to maintain access to putative regulatory regions throughout all stages of development may be responsible for plants' incredible ability to regenerate body structures, especially in comparison to most animals. It would be illuminating to examine the epigenomes of exceptionally regenerative animals, such as starfish, and determine whether they share any epigenetic similarities with plant species, such as persistent accessibility of regulatory elements. Also, it would be of interest to discover whether these similarities are constant throughout the lifetime of the animal, or are instead induced during the regenerative process, the latter of which may present an opportunity to someday adapt these epigenetic mechanisms for use in wound healing in humans.

We also performed ChIP-seq experiments in *Arabidopsis* for the most common enhancer histone post-translational modifications (PTMs) with single cell-type specificity (**Chapter 4**). Upon examining PTM enrichment at promoters and gene bodies, it was clear that fundamental differences exist between the model plant and model animals, even from a global perspective. Where human and *Drosophila* plots exhibit a bimodal peak, characteristic of the deposition of histone marks during the processes of sense- and antisense-transcription, *Arabidopsis* plots show an exclusive bias for sense-transcription. This pattern of sense-transcription is recapitulated at accessible chromatin sites/putative regulatory regions across the genome, and is further confirmed by nascent transcriptional data.

What could be responsible for this stark contrast between transcriptional direction in *Arabidopsis* and model animal species? Recent findings from Hi-C data with single-gene resolution may shed some light on this 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 (Liu, Wang et al. 2016). The constrained geometry of these gene loops has been shown to eliminate bidirectional transcription, forcing RNAPII to transcribe in the sense

direction alone (Tan-Wong, Zaugg et al. 2012). This organizational scheme and lack of long-range interactions could explain why the overwhelming majority of accessible chromatin sties we identified were preferentially located proximally upstream of their nearest gene. A subunit of RNAPII, 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. While it is not yet known whether plants contain a functional ortholog to Ssu72, these findings suggest that differences in higher order chromatin structure may be responsible for distinctions in transcriptional direction. Additionally, the apparent absence of 5' RNAPII pausing in *Arabidopsis* and maize (Hetzel, Duttke et al. 2016), in conjunction with the lack of negative elongation factor (NELF) in plants (Wu, Yamaguchi et al. 2003) further supports that transcriptional regulation in plants has notable distinctions from the mechanisms in animals, and is likely more heavily regulated at the level of initiation.

These trends strengthen the finding that the classic epigenetic enhancer profile that has been accumulated from animal studies is not universal amongst complex, multicellular eukaryotes. Beyond our own research, other recent studies have shed light regarding histone modification enrichment at *cis*-regulatory regions in plant species at large. Though animal enhancer-associated modifications such as H3K9ac and H3K27ac have been found at accessible chromatin regions (Zhu, Zhang et al. 2015, Oka, Zicola et al. 2017), maize studies have revealed that H3K4me1 is not present at putative enhancers (Oka, Zicola et al. 2017, Ricci, Lu et al. 2019). As H3K4me1 is the histone modification conserved at enhancers of all activity states in metazoans, this may belie a fundamental divergence between plants and animals. Notably, a recent study performed a massive comparison between the chromatin accessibility, enrichment for five histone PTMs, and sequence conservation of putative regulatory regions in thirteen angiosperm species (Lu, Marand et al. 2019). Though this investigation used whole leaf and whole seedling tissue as input for their sequencing libraries, their findings align well with the results of our single cell type dataset analyses. Using ATAC-seq, this study identified accessible chromatin sites across the genomes of interest, the majority of which were within or proximal to genes. The relative amount of distal accessible chromatin regions (dACRs) varied greatly from species to species, ranging from 5% to 40% and increasing

dramatically with increasing genome size. Within these distal accessible chromatin regions (dACRs), the authors found four distinct chromatin states, none of which match the enhancer signature characterized in animal species. The states include 'unmodified', 'Kac' (H3K56ac), 'K27me3' (H3K27me3), and 'transcribed' (H3K4me3, H3K56ac, H3K36me3), which match the chromatin states identified in a recent deep-dive chromatin study in maize (Ricci, Lu et al. 2019).

While it is tempting to make the conclusion that the transcriptional machinery in plant species fundamentally differs from that used in animals, as evidenced by this difference in histone modifications, it is important to keep in mind the diversity that has already been uncovered in animals. In studies using non-methylatable H3, H3K4me3 was found not to be necessary for transcription in *Drosophila* (Hodl and Basler 2009, Hodl and Basler 2012). Similarly, the enhancer modification H3K4me1 has been found to be dispensable for enhancer activity (Dorighi, Swigut et al. 2017, Rickels, Herz et al. 2017). As such, it is important to be mindful of the fact that correlation does not imply causation, and that while these histone PTMs are often present, they may not be crucial to the function of the element itself (Pollex and Furlong 2017). Thus, basing our search for novel plant enhancers based on these potentially dispensable criteria inherently involves risk.

Future Directions – Plant Enhancer Discovery with STARR-seq

The future of *cis*-regulatory element discovery and characterization depends on using approaches which do not rely on the historical definitions of histone PTM enrichment, genomic location, or inability to initiate transcription. Asking 'does this candidate sequence meet this set of physical criteria?', the boundaries of which are ever-changing, is no longer sufficient to identify elements and reliably differentiate between regulatory classes. Instead, secondary criteria must be stripped away to allow candidate enhancers to be screened based on their most fundamental identifier: the ability to amplify transcriptional output of a promoter.

The most promising new approach that can assess elements based on functional metrics is the method of Self-Transcribing Active Regulatory Regions-sequencing (STARR-seq) (Arnold, Gerlach et al.

2013, Shlyueva, Stampfel et al. 2014, Muerdter, Boryn et al. 2015). Distinct from other high-throughput sequencing assays, STARR-seq identifies enhancers based solely on functional output, independent of the sequence's location, orientation, or secondary epigenetic characteristics. It has been used with great success in human ESCs (Barakat, Halbritter et al. 2018) and immortalized cell lines (Liu, Liu et al. 2017, Liu, Yu et al. 2017, Klein, Keith et al. 2018, Muerdter, Boryn et al. 2018, Zhang, Xia et al. 2018), *Drosophila* cell lines (Arnold, Gerlach et al. 2014, Cubenas-Potts, Rowley et al. 2017), and has even been adapted for use in a synthetic system (Schone, Bothe et al. 2018). However, STARR-seq has yet to be applied to an organism outside of the animal kingdom. It was the goal of my United States Department of Agriculture (USDA) National Institute of Food and Agriculture (NIFA) predoctoral fellowship and the continuing goal of a collaborative project with the University of Washington to adapt STARR-seq for use in plant enhancer discovery.

The workflow of STARR-seq is outlined in **Figure 5.1**. First, genomic DNA fragments are cloned into minimal promoter-reporter constructs, and then transfected into a transient expression system. In many plants, including *Arabidopsis*, transient expression is most readily achievable by transfection of protoplasts, plant cells whose cell wall has been enzymatically digested, leaving the underlying cellular membrane intact (Mathur and Koncz 1998). After transfection, any DNA fragments with heterologous enhancer activity will be able to stimulate the construct's upstream minimal promoter, causing a fusion transcript consisting of both the reporter and the DNA fragment sequences to be generated. STARR-seq vector-derived transcripts are then selectively extracted from the transient expression system, sequenced, and all putative enhancers can then be identified by their sequences in the fusion transcripts. In this way, STARR-seq is able to analyze a large pool of candidates in a high-throughput manner based on stringent functional criteria. Furthermore, STARR-seq has been demonstrated to be able to identify enhancers in heterochromatin (Arnold, Gerlach et al. 2013), ensuring that a wide range of potential enhancers can be interrogated regardless of their original location within the epigenome.

Figure 5.2 shows the results of my work towards my NIFA fellowship. After much optimization, I had success in both isolating and transforming *Arabidopsis* mesophyll protoplasts, both with a standard

GFP reporter construct (**Figure 5.2A**) and with a genuine STARR-seq library generated from fragments covering the whole *Arabidopsis* genome. **Figure 5.2B** shows the results of a pilot experiment with a small input library (1.1 million *Arabidopsis* genome fragments) in *Arabidopsis* protoplasts and the overlap between the putative enhancer fragments isolated from two biological replicates. Fragments that are highly enriched in the STARR-seq output (left panel, blue) represent potential candidates for strong enhancers. These same candidate fragment sequences are found to be highly enriched in both biological replicates (right panel, blue box), demonstrating that the results of this experiment are highly reproducible.

Further optimization of the *Arabidopsis* STARR-seq library continues, specifically with perfecting the minimal promoter in the reporter library construct, with the aim of identifying moderate and weak putative enhancers in the future. Simultaneously with this project, I was also involved with a collaborative project with Christine Queitsch's lab at the University of Washington. Together with Edward Buckler from Cornell University, Jorge Dubcovsky from the University of California Davis, and Stanley Fields from the University of Washington, we received funding from the National Science Foundation (NSF) to interrogate transcriptional regulation in key crop species. The aim of the grant that I was most directly involved in was the adaptation of STARR-seq for use in rice, sorghum, and maize – three of the most prominent crop species around the globe. These crops presented a challenge that *Arabidopsis* did not; in addition to having much larger genomes (375 Mb, 730 Mb, and 2,300 Mb respectively), these species also contained large regions of genomic repeats, which often confound the mapping of sequencing reads. As a result, it would be extraordinarily difficult to gain sufficient sequencing coverage if the entire genome was used as the library input, as we had done previously with *Arabidopsis*.

Knowing that active *cis*-regulatory elements exist preferentially in accessible chromatin, a finding that was borne out in our own studies (Maher, Bajic et al. 2018), an ingenious approach to reduce the number of input fragments was devised. To decrease the amount of genomic noise in our target list, an ATAC-seq library was first created for each species and then utilized as the input for the creation of a STARR-seq library. ATAC-seq and STARR-seq have been successfully combined before in human cell lines(Wang, He et al. 2018), and the preliminary results in crop species have been promising in our hands.

Though there is certainly further optimization to be done, these initial results demonstrate the feasibility of this line of experimental investigation. These data, especially when analyzed in combination with the existing information on chromatin architecture and interacting domains (Dong, Tu et al. 2017) may give us a better idea of what transcriptional regulation looks like both in *Arabidopsis* and in plants systems at large.

It has been the goal of this dissertation to both develop and implement a variety of next-generation sequencing approaches to characterize putative enhancers in plant species. Through the creation of INTACT-ATAC-seq, its adaptation to *Arabidopsis*, *Medicago*, rice, and tomato, and the application of single cell-type ChIP-seq to map enhancer-specific histone modification marks across the genome, we have interrogated the nature of genetic enhancers in plants based on multiple criteria. Through reexamining the fundamentals of the accepted definitions of enhancers and promoters and by working to adapt STARR-seq for use in model plant and vital crop species, we have the opportunity to pursue plant enhancers on their own terms, and hope to open up unprecedented avenues of investigation into transcriptional control.

FIGURES

Note: All diagrams and microscopy images in this section are original works and were created by the author, Kelsey A. Maher. STARR-seq data was generated by Kelsey A. Maher in cooperation with collaborators in the Queitsch lab at the University of Washington.

Figure 5.1

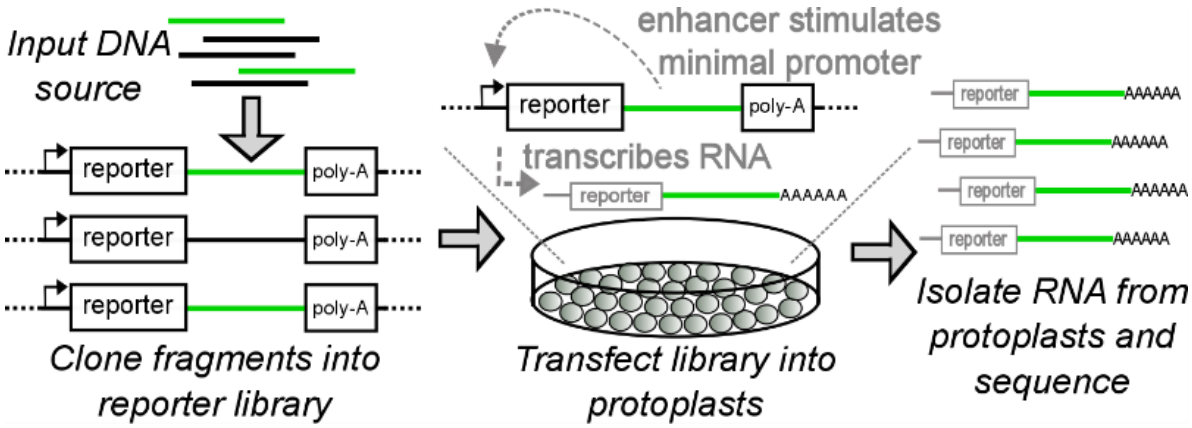


Figure 5.1: STARR-seq is a powerful tool for high-throughput enhancer identification. Self-Transcribing Active Regulatory Regions with high-throughput sequencing (STARR-seq) begins with a source of DNA, which is fragmented and cloned into reporter constructs made of a minimal promoter, a reporter ORF (such as GFP), and a polyadenylation sequence. Fragments containing an enhancer element are shown as green lines; fragments without an enhancer are black lines. Mesophyll protoplasts are transfected with the reporter library. In any reporter construct with a fragment containing an enhancer, the enhancer will stimulate the minimal promoter, causing transcription of a reporter-enhancer fusion transcript. RNA is purified from the protoplasts and sequenced. Transcripts are distinguished from endogenous transcripts by the reporter sequence, and distinguished from other enhancer fusion transcripts by the enhancer sequence itself.

Figure 5.2

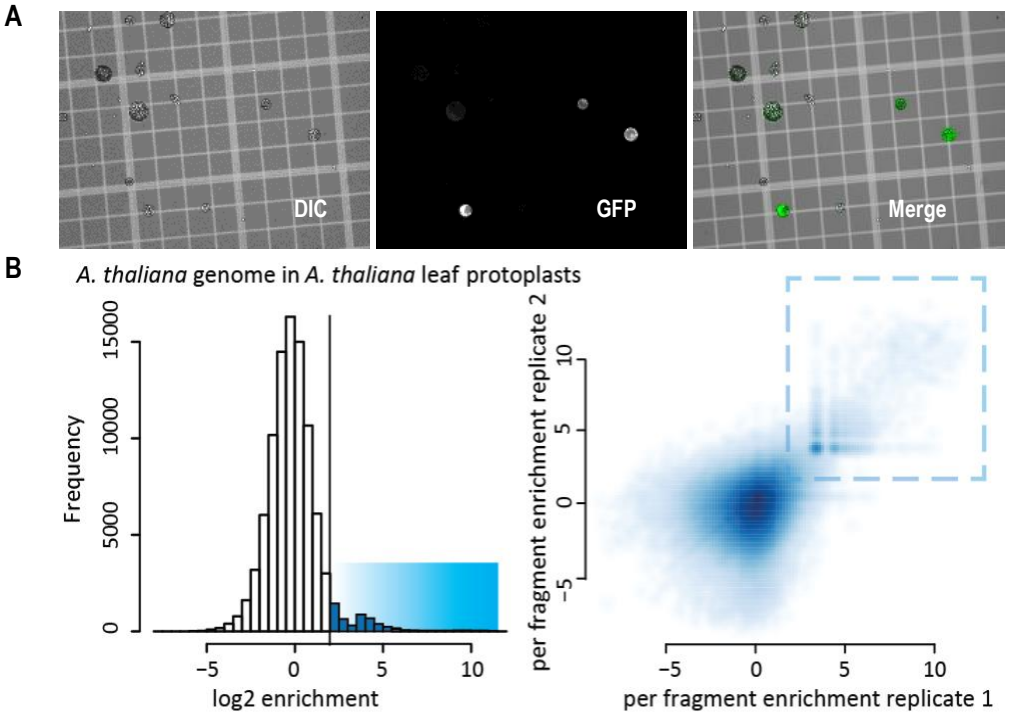


Figure 5.2: STARR-seq shows promising and reproducible results. (A) *Arabidopsis* mesophyll protoplasts transformed with a test GFP reporter plasmid (pHBT-sGFP(S65T)-NOS). Transformation conditions included 40% polyethylene glycol (PEG) with 20 μg of DNA for a 15-minute incubation. (**B)** A pilot run of an *Arabidopsis* genome STARR-seq library expressed in *Arabidopsis* mesophyll protoplasts. Many fragments were isolated that were highly active as enhancers (left plot, in blue). The boxed region in the right panel shows that highly enriched fragments are highly reproducibility between biological replicates.

LITERATURE CITED

(2004). "The ENCODE (ENCyclopedia Of DNA Elements) Project." Science **306**(5696): 636-640.

(2012). "An integrated encyclopedia of DNA elements in the human genome." Nature **489**(7414): 57-74.

Adli, M. and B. E. Bernstein (2011). "Whole-genome chromatin profiling from limited numbers of cells using nano-ChIP-seq." Nat Protoc **6**(10): 1656-1668.

Allshire, R. C. and K. Ekwall (2015). "Epigenetic Regulation of Chromatin States in Schizosaccharomyces pombe." Cold Spring Harb Perspect Biol **7**(7): a018770.

Amendola, M. and B. van Steensel (2014). "Mechanisms and dynamics of nuclear lamina-genome interactions." Curr Opin Cell Biol **28**: 61-68.

Anders, S. and W. Huber (2010). "Differential expression analysis for sequence count data." Genome Biol **11**(10): R106.

Andersson, R., C. Gebhard, I. Miguel-Escalada, I. Hoof, J. Bornholdt, M. Boyd, Y. Chen, X. Zhao, C. Schmidl, T. Suzuki, E. Ntini, E. Arner, E. Valen, K. Li, L. Schwarzfischer, D. Glatz, J. Raithel, B. Lilje, N. Rapin, F. O. Bagger, M. Jørgensen, P. R. Andersen, N. Bertin, O. Rackham, A. M. Burroughs, J. K. Baillie, Y. Ishizu, Y. Shimizu, E. Furuhata, S. Maeda, Y. Negishi, C. J. Mungall, T. F. Meehan, T. Lassmann, M. Itoh, H. Kawaji, N. Kondo, J. Kawai, A. Lennartsson, C. O. Daub, P. Heutink, D. A. Hume, T. H. Jensen, H. Suzuki, Y. Hayashizaki, F. Müller, A. R. R. Forrest, P. Carninci, M. Rehli and A. Sandelin (2014). "An atlas of active enhancers across human cell types and tissues." Nature **507**(7493): 455-461.

Angeloni, A. and O. Bogdanovic (2019). "Enhancer DNA methylation: implications for gene regulation." Essays Biochem **63**(6): 707-715.

Arents, G., R. W. Burlingame, B. C. Wang, W. E. Love and E. N. Moudrianakis (1991). "The nucleosomal core histone octamer at 3.1 A resolution: a tripartite protein assembly and a left-handed superhelix." Proc Natl Acad Sci U S A **88**(22): 10148-10152.

Arnold, C. D., D. Gerlach, D. Spies, J. A. Matts, Y. A. Sytnikova, M. Pagani, N. C. Lau and A. Stark (2014). "Quantitative genome-wide enhancer activity maps for five Drosophila species show functional enhancer conservation and turnover during cis-regulatory evolution." Nat Genet **46**(7): 685-692.

Arnold, C. D., D. Gerlach, C. Stelzer, L. M. Boryn, M. Rath and A. Stark (2013). "Genome-wide quantitative enhancer activity maps identified by STARR-seq." Science **339**(6123): 1074-1077.

Arnold, C. D., M. A. Zabidi, M. Pagani, M. Rath, K. Schernhuber, T. Kazmar and A. Stark (2017). "Genome-wide assessment of sequence-intrinsic enhancer responsiveness at single-base-pair resolution." Nat Biotechnol **35**(2): 136-144.

Atchison, M. L. (1988). "Enhancers: mechanisms of action and cell specificity." Annu Rev Cell Biol **4**: 127-153.

Ayer, S. and C. Benyajati (1990). "Conserved enhancer and silencer elements responsible for differential Adh transcription in Drosophila cell lines." Mol Cell Biol **10**(7): 3512-3523.

Backstrom, S., N. Elfving, R. Nilsson, G. Wingsle and S. Bjorklund (2007). "Purification of a plant mediator from Arabidopsis thaliana identifies PFT1 as the Med25 subunit." Mol Cell **26**(5): 717-729.

Banerji, J., L. Olson and W. Schaffner (1983). "A lymphocyte-specific cellular enhancer is located downstream of the joining region in immunoglobulin heavy chain genes." Cell **33**(3): 729-740.

Banerji, J., S. Rusconi and W. Schaffner (1981). "Expression of a beta-globin gene is enhanced by remote SV40 DNA sequences." Cell **27**(2 Pt 1): 299-308.

Barakat, T. S., F. Halbritter, M. Zhang, A. F. Rendeiro, E. Perenthaler, C. Bock and I. Chambers (2018). "Functional Dissection of the Enhancer Repertoire in Human Embryonic Stem Cells." Cell Stem Cell **23**(2): 276-288.e278.

Barski, A., S. Cuddapah, K. Cui, T. Y. Roh, D. E. Schones, Z. Wang, G. Wei, I. Chepelev and K. Zhao (2007). "High-resolution profiling of histone methylations in the human genome." Cell **129**(4): 823-837.

Bell, O., V. K. Tiwari, N. H. Thoma and D. Schubeler (2011). "Determinants and dynamics of genome accessibility." Nat Rev Genet **12**(8): 554-564.

Benoist, C. and P. Chambon (1981). "In vivo sequence requirements of the SV40 early promotor region." Nature **290**(5804): 304-310.

Berg, P. E., Z. Popovic and W. F. Anderson (1984). "Promoter dependence of enhancer activity." Mol Cell Biol **4**(8): 1664-1668.

Bernstein, B. E., T. S. Mikkelsen, X. Xie, M. Kamal, D. J. Huebert, J. Cuff, B. Fry, A. Meissner, M. Wernig, K. Plath, R. Jaenisch, A. Wagschal, R. Feil, S. L. Schreiber and E. S. Lander (2006). "A bivalent chromatin structure marks key developmental genes in embryonic stem cells." Cell **125**(2): 315-326.

Birney, E., J. A. Stamatoyannopoulos, A. Dutta, R. Guigó, T. R. Gingeras, E. H. Margulies, Z. Weng, M. Snyder, E. T. Dermitzakis, R. E. Thurman, M. S. Kuehn, C. M. Taylor, S. Neph, C. M. Koch, S. Asthana, A. Malhotra, I. Adzhubei, J. A. Greenbaum, R. M. Andrews, P. Flicek, P. J. Boyle, H. Cao, N. P. Carter, G. K. Clelland, S. Davis, N. Day, P. Dhami, S. C. Dillon, M. O. Dorschner, H. Fiegler, P. G. Giresi, J. Goldy, M. Hawrylycz, A. Haydock, R. Humbert, K. D. James, B. E. Johnson, E. M. Johnson, T. T. Frum, E. R. Rosenzweig, N. Karnani, K. Lee, G. C. Lefebvre, P. A. Navas, F. Neri, S. C. Parker, P. J. Sabo, R. Sandstrom, A. Shafer, D. Vetrie, M. Weaver, S. Wilcox, M. Yu, F. S. Collins, J. Dekker, J. D. Lieb, T. D. Tullius, G. E. Crawford, S. Sunyaev, W. S. Noble, I. Dunham, F. Denoeud, A. Reymond, P. Kapranov, J. Rozowsky, D. Zheng, R. Castelo, A. Frankish, J. Harrow, S. Ghosh, A. Sandelin, I. L. Hofacker, R. Baertsch, D. Keefe, S. Dike, J. Cheng, H. A. Hirsch, E. A. Sekinger, J. Lagarde, J. F. Abril, A. Shahab, C. Flamm, C. Fried, J. Hackermüller, J. Hertel, M. Lindemeyer, K. Missal, A. Tanzer, S. Washietl, J. Korbel, O. Emanuelsson, J. S. Pedersen, N. Holroyd, R. Taylor, D. Swarbreck, N. Matthews, M. C. Dickson, D. J. Thomas, M. T. Weirauch, J. Gilbert, J. Drenkow, I. Bell, X. Zhao, K. G. Srinivasan, W. K. Sung, H. S. Ooi, K. P. Chiu, S. Foissac, T. Alioto, M. Brent, L. Pachter, M. L. Tress, A. Valencia, S. W. Choo, C. Y. Choo, C. Ucla, C. Manzano, C. Wyss, E. Cheung, T. G. Clark, J. B. Brown, M. Ganesh, S. Patel, H. Tammana, J.

Chrast, C. N. Henrichsen, C. Kai, J. Kawai, U. Nagalakshmi, J. Wu, Z. Lian, J. Lian, P. Newburger, X. Zhang, P. Bickel, J. S. Mattick, P. Carninci, Y. Hayashizaki, S. Weissman, T. Hubbard, R. M. Myers, J. Rogers, P. F. Stadler, T. M. Lowe, C. L. Wei, Y. Ruan, K. Struhl, M. Gerstein, S. E. Antonarakis, Y. Fu, E. D. Green, U. Karaöz, A. Siepel, J. Taylor, L. A. Liefer, K. A. Wetterstrand, P. J. Good, E. A. Feingold, M. S. Guyer, G. M. Cooper, G. Asimenos, C. N. Dewey, M. Hou, S. Nikolaev, J. I. Montoya-Burgos, A. Löytynoja, S. Whelan, F. Pardi, T. Massingham, H. Huang, N. R. Zhang, I. Holmes, J. C. Mullikin, A. Ureta-Vidal, B. Paten, M. Seringhaus, D. Church, K. Rosenbloom, W. J. Kent, E. A. Stone, S. Batzoglou, N. Goldman, R. C. Hardison, D. Haussler, W. Miller, A. Sidow, N. D. Trinklein, Z. D. Zhang, L. Barrera, R. Stuart, D. C. King, A. Ameur, S. Enroth, M. C. Bieda, J. Kim, A. A. Bhinge, N. Jiang, J. Liu, F. Yao, V. B. Vega, C. W. Lee, P. Ng, A. Shahab, A. Yang, Z. Moqtaderi, Z. Zhu, X. Xu, S. Squazzo, M. J. Oberley, D. Inman, M. A. Singer, T. A. Richmond, K. J. Munn, A. Rada-Iglesias, O. Wallerman, J. Komorowski, J. C. Fowler, P. Couttet, A. W. Bruce, O. M. Dovey, P. D. Ellis, C. F. Langford, D. A. Nix, G. Euskirchen, S. Hartman, A. E. Urban, P. Kraus, S. Van Calcar, N. Heintzman, T. H. Kim, K. Wang, C. Qu, G. Hon, R. Luna, C. K. Glass, M. G. Rosenfeld, S. F. Aldred, S. J. Cooper, A. Halees, J. M. Lin, H. P. Shulha, X. Zhang, M. Xu, J. N. Haidar, Y. Yu, Y. Ruan, V. R. Iyer, R. D. Green, C. Wadelius, P. J. Farnham, B. Ren, R. A. Harte, A. S. Hinrichs, H. Trumbower, H. Clawson, J. Hillman-Jackson, A. S. Zweig, K. Smith, A. Thakkapallayil, G. Barber, R. M. Kuhn, D. Karolchik, L. Armengol, C. P. Bird, P. I. de Bakker, A. D. Kern, N. Lopez-Bigas, J. D. Martin, B. E. Stranger, A. Woodroffe, E. Davydov, A. Dimas, E. Eyras, I. B. Hallgrímsdóttir, J. Huppert, M. C. Zody, G. R. Abecasis, X. Estivill, G. G. Bouffard, X. Guan, N. F. Hansen, J. R. Idol, V. V. Maduro, B. Maskeri, J. C. McDowell, M. Park, P. J. Thomas, A. C. Young, R. W. Blakesley, D. M. Muzny, E. Sodergren, D. A. Wheeler, K. C. Worley, H. Jiang, G. M. Weinstock, R. A. Gibbs, T. Graves, R. Fulton, E. R. Mardis, R. K. Wilson, M. Clamp, J. Cuff, S. Gnerre, D. B. Jaffe, J. L. Chang, K. Lindblad-Toh, E. S. Lander, M. Koriabine, M. Nefedov, K. Osoegawa, Y. Yoshinaga, B. Zhu and P. J. de Jong (2007). "Identification and

analysis of functional elements in 1% of the human genome by the ENCODE pilot project." Nature **447**(7146): 799-816.

Bolduc, N., A. Yilmaz, M. K. Mejia-Guerra, K. Morohashi, D. O'Connor, E. Grotewold and S. Hake (2012). "Unraveling the KNOTTED1 regulatory network in maize meristems." Genes Dev **26**(15): 1685-1690.

Boley, N., K. H. Wan, P. J. Bickel and S. E. Celniker (2014). "Navigating and mining modENCODE data." Methods **68**(1): 38-47.

Bonn, S., R. P. Zinzen, C. Girardot, E. H. Gustafson, A. Perez-Gonzalez, N. Delhomme, Y. Ghavi-Helm, B. Wilczynski, A. Riddell and E. E. Furlong (2012). "Tissue-specific analysis of chromatin state identifies temporal signatures of enhancer activity during embryonic development." Nat Genet **44**(2): 148-156.

Bonn, S., R. P. Zinzen, A. Perez-Gonzalez, A. Riddell, A. C. Gavin and E. E. Furlong (2012). "Cell type-specific chromatin immunoprecipitation from multicellular complex samples using BiTS-ChIP." Nat Protoc **7**(5): 978-994.

Boyle, A. P., S. Davis, H. P. Shulha, P. Meltzer, E. H. Margulies, Z. Weng, T. S. Furey and G. E. Crawford (2008). "High-resolution mapping and characterization of open chromatin across the genome." Cell **132**(2): 311-322.

Brangwynne, C. P., C. R. Eckmann, D. S. Courson, A. Rybarska, C. Hoege, J. Gharakhani, F. Julicher and A. A. Hyman (2009). "Germline P granules are liquid droplets that localize by controlled dissolution/condensation." Science **324**(5935): 1729-1732.

Buenrostro, J. D., P. G. Giresi, L. C. Zaba, H. Y. Chang and W. J. Greenleaf (2013). "Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position." Nat Methods **10**(12): 1213-1218.

Buenrostro, J. D., B. Wu, H. Y. Chang and W. J. Greenleaf (2015). "ATAC-seq: A Method for Assaying Chromatin Accessibility Genome-Wide." Curr Protoc Mol Biol **109**: 21.29.21-29.

Buhler, M. and S. M. Gasser (2009). "Silent chromatin at the middle and ends: lessons from yeasts." Embo j **28**(15): 2149-2161.

Burke, T. W. and J. T. Kadonaga (1996). "Drosophila TFIID binds to a conserved downstream basal promoter element that is present in many TATA-box-deficient promoters." Genes Dev **10**(6): 711-724.

Canzio, D., E. Y. Chang, S. Shankar, K. M. Kuchenbecker, M. D. Simon, H. D. Madhani, G. J. Narlikar and B. Al-Sady (2011). "Chromodomain-mediated oligomerization of HP1 suggests a nucleosome-bridging mechanism for heterochromatin assembly." Mol Cell **41**(1): 67-81.

Carter, D., L. Chakalova, C. S. Osborne, Y. F. Dai and P. Fraser (2002). "Long-range chromatin regulatory interactions in vivo." Nat Genet **32**(4): 623-626.

Chatterjee, S. and N. Ahituv (2017). "Gene Regulatory Elements, Major Drivers of Human Disease." Annu Rev Genomics Hum Genet **18**: 45-63.

Chudalayandi, S. (2011). "Enhancer trapping in plants." Methods Mol Biol **701**: 285-300.

Clapier, C. R. and B. R. Cairns (2009). "The biology of chromatin remodeling complexes." Annu Rev Biochem **78**: 273-304.

Clark, R. M., T. N. Wagler, P. Quijada and J. Doebley (2006). "A distant upstream enhancer at the maize domestication gene tb1 has pleiotropic effects on plant and inflorescent architecture." Nat Genet **38**(5): 594-597.

Core, L. J., A. L. Martins, C. G. Danko, C. T. Waters, A. Siepel and J. T. Lis (2014). "Analysis of nascent RNA identifies a unified architecture of initiation regions at mammalian promoters and enhancers." Nat Genet **46**(12): 1311-1320.

Core, L. J., J. J. Waterfall and J. T. Lis (2008). "Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters." Science **322**(5909): 1845-1848.

Creyghton, M. P., A. W. Cheng, G. G. Welstead, T. Kooistra, B. W. Carey, E. J. Steine, J. Hanna, M. A. Lodato, G. M. Frampton, P. A. Sharp, L. A. Boyer, R. A. Young and R. Jaenisch (2010). "Histone

H3K27ac separates active from poised enhancers and predicts developmental state." Proc Natl Acad Sci U S A **107**(50): 21931-21936.

Cubenas-Potts, C., M. J. Rowley, X. Lyu, G. Li, E. P. Lei and V. G. Corces (2017). "Different enhancer classes in Drosophila bind distinct architectural proteins and mediate unique chromatin interactions and 3D architecture." Nucleic Acids Res **45**(4): 1714-1730.

Dao, L. T. M., A. O. Galindo-Albarran, J. A. Castro-Mondragon, C. Andrieu-Soler, A. Medina-Rivera, C. Souaid, G. Charbonnier, A. Griffon, L. Vanhille, T. Stephen, J. Alomairi, D. Martin, M. Torres, N. Fernandez, E. Soler, J. van Helden, D. Puthier and S. Spicuglia (2017). "Genome-wide characterization of mammalian promoters with distal enhancer functions." Nat Genet **49**(7): 1073-1081.

Davis, C. A., B. C. Hitz, C. A. Sloan, E. T. Chan, J. M. Davidson, I. Gabdank, J. A. Hilton, K. Jain, U. K. Baymuradov, A. K. Narayanan, K. C. Onate, K. Graham, S. R. Miyasato, T. R. Dreszer, J. S. Strattan, O. Jolanki, F. Y. Tanaka and J. M. Cherry (2018). "The Encyclopedia of DNA elements (ENCODE): data portal update." Nucleic Acids Res **46**(D1): D794-d801.

De Santa, F., I. Barozzi, F. Mietton, S. Ghisletti, S. Polletti, B. K. Tusi, H. Muller, J. Ragoussis, C. L. Wei and G. Natoli (2010). "A large fraction of extragenic RNA pol II transcription sites overlap enhancers." PLoS Biol **8**(5): e1000384.

Deal, R. B. and S. Henikoff (2010). "A simple method for gene expression and chromatin profiling of individual cell types within a tissue." Dev Cell **18**(6): 1030-1040.

Derrien, T., R. Johnson, G. Bussotti, A. Tanzer, S. Djebali, H. Tilgner, G. Guernec, D. Martin, A. Merkel, D. G. Knowles, J. Lagarde, L. Veeravalli, X. Ruan, Y. Ruan, T. Lassmann, P. Carninci, J. B. Brown, L. Lipovich, J. M. Gonzalez, M. Thomas, C. A. Davis, R. Shiekhattar, T. R. Gingeras, T. J. Hubbard, C. Notredame, J. Harrow and R. Guigo (2012). "The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression." Genome Res **22**(9): 1775-1789.

Djebali, S., C. A. Davis, A. Merkel, A. Dobin, T. Lassmann, A. Mortazavi, A. Tanzer, J. Lagarde, W. Lin, F. Schlesinger, C. Xue, G. K. Marinov, J. Khatun, B. A. Williams, C. Zaleski, J. Rozowsky, M. Roder, F. Kokocinski, R. F. Abdelhamid, T. Alioto, I. Antoshechkin, M. T. Baer, N. S. Bar, P. Batut, K. Bell, I. Bell, S. Chakrabortty, X. Chen, J. Chrast, J. Curado, T. Derrien, J. Drenkow, E. Dumais, J. Dumais, R. Duttagupta, E. Falconnet, M. Fastuca, K. Fejes-Toth, P. Ferreira, S. Foissac, M. J. Fullwood, H. Gao, D. Gonzalez, A. Gordon, H. Gunawardena, C. Howald, S. Jha, R. Johnson, P. Kapranov, B. King, C. Kingswood, O. J. Luo, E. Park, K. Persaud, J. B. Preall, P. Ribeca, B. Risk, D. Robyr, M. Sammeth, L. Schaffer, L. H. See, A. Shahab, J. Skancke, A. M. Suzuki, H. Takahashi, H. Tilgner, D. Trout, N. Walters, H. Wang, J. Wrobel, Y. Yu, X. Ruan, Y. Hayashizaki, J. Harrow, M. Gerstein, T. Hubbard, A. Reymond, S. E. Antonarakis, G. Hannon, M. C. Giddings, Y. Ruan, B. Wold, P. Carninci, R. Guigo and T. R. Gingeras (2012). "Landscape of transcription in human cells." Nature **489**(7414): 101-108.

Dong, P., X. Tu, P. Y. Chu, P. Lu, N. Zhu, D. Grierson, B. Du, P. Li and S. Zhong (2017). "3D Chromatin Architecture of Large Plant Genomes Determined by Local A/B Compartments." Mol Plant **10**(12): 1497-1509.

Dorighi, K. M., T. Swigut, T. Henriques, N. V. Bhanu, B. S. Scruggs, N. Nady, C. D. Still, 2nd, B. A. Garcia, K. Adelman and J. Wysocka (2017). "Mll3 and Mll4 Facilitate Enhancer RNA Synthesis and Transcription from Promoters Independently of H3K4 Monomethylation." Mol Cell **66**(4): 568-576.e564.

Ebmeier, C. C. and D. J. Taatjes (2010). "Activator-Mediator binding regulates Mediator-cofactor interactions." Proc Natl Acad Sci U S A **107**(25): 11283-11288.

Edgar, R., M. Domrachev and A. E. Lash (2002). "Gene Expression Omnibus: NCBI gene expression and hybridization array data repository." Nucleic Acids Res **30**(1): 207-210.

Ernst, J. and M. Kellis (2010). "Discovery and characterization of chromatin states for systematic annotation of the human genome." Nat Biotechnol **28**(8): 817-825.

Ernst, J., P. Kheradpour, T. S. Mikkelsen, N. Shoresh, L. D. Ward, C. B. Epstein, X. Zhang, L. Wang, R. Issner, M. Coyne, M. Ku, T. Durham, M. Kellis and B. E. Bernstein (2011). "Mapping and analysis of chromatin state dynamics in nine human cell types." Nature **473**(7345): 43-49.

Faure, A. J., D. Schmidt, S. Watt, P. C. Schwalie, M. D. Wilson, H. Xu, R. G. Ramsay, D. T. Odom and P. Flicek (2012). "Cohesin regulates tissue-specific expression by stabilizing highly occupied cis-regulatory modules." Genome Res **22**(11): 2163-2175.

Feng, S., S. J. Cokus, V. Schubert, J. Zhai, M. Pellegrini and S. E. Jacobsen (2014). "Genome-wide Hi-C analyses in wild-type and mutants reveal high-resolution chromatin interactions in Arabidopsis." Mol Cell **55**(5): 694-707.

FitzGerald, P. C., D. Sturgill, A. Shyakhtenko, B. Oliver and C. Vinson (2006). "Comparative genomics of Drosophila and human core promoters." Genome Biol **7**(7): R53.

Francis, N. J., R. E. Kingston and C. L. Woodcock (2004). "Chromatin compaction by a polycomb group protein complex." Science **306**(5701): 1574-1577.

Fromm, M. and P. Berg (1982). "Deletion mapping of DNA regions required for SV40 early region promoter function in vivo." J Mol Appl Genet **1**(5): 457-481.

Fromm, M. and P. Berg (1983). "Simian virus 40 early- and late-region promoter functions are enhanced by the 72-base-pair repeat inserted at distant locations and inverted orientations." Mol Cell Biol **3**(6): 991-999.

Furner, I. J. and M. Matzke (2011). "Methylation and demethylation of the Arabidopsis genome." Curr Opin Plant Biol **14**(2): 137-141.

Gardiner-Garden, M. and M. Frommer (1987). "CpG islands in vertebrate genomes." J Mol Biol **196**(2): 261-282.

Gardini, A. (2017). "Global Run-On Sequencing (GRO-Seq)." Methods Mol Biol **1468**: 111-120.

Gaszner, M. and G. Felsenfeld (2006). "Insulators: exploiting transcriptional and epigenetic mechanisms." Nat Rev Genet **7**(9): 703-713.

Ghavi-Helm, Y., F. A. Klein, T. Pakozdi, L. Ciglar, D. Noordermeer, W. Huber and E. E. Furlong (2014). "Enhancer loops appear stable during development and are associated with paused polymerase." Nature **512**(7512): 96-100.

Gilchrist, D. A., G. Dos Santos, D. C. Fargo, B. Xie, Y. Gao, L. Li and K. Adelman (2010). "Pausing of RNA polymerase II disrupts DNA-specified nucleosome organization to enable precise gene regulation." Cell **143**(4): 540-551.

Gillies, S. D., S. L. Morrison, V. T. Oi and S. Tonegawa (1983). "A tissue-specific transcription enhancer element is located in the major intron of a rearranged immunoglobulin heavy chain gene." Cell **33**(3): 717-728.

Gonzalez-Sandoval, A. and S. M. Gasser (2016). "On TADs and LADs: Spatial Control Over Gene Expression." Trends Genet **32**(8): 485-495.

Grant, C. E., T. L. Bailey and W. S. Noble (2011). "FIMO: scanning for occurrences of a given motif." Bioinformatics **27**(7): 1017-1018.

Green, P. J., S. A. Kay and N. H. Chua (1987). "Sequence-specific interactions of a pea nuclear factor with light-responsive elements upstream of the rbcS-3A gene." Embo j **6**(9): 2543-2549.

Grosschedl, R. and M. L. Birnstiel (1980). "Spacer DNA sequences upstream of the T-A-T-A-A-A-T-A sequence are essential for promotion of H2A histone gene transcription in vivo." Proc Natl Acad Sci U S A **77**(12): 7102-7106.

Grossniklaus, U. and R. Paro (2014). "Transcriptional silencing by polycomb-group proteins." Cold Spring Harb Perspect Biol **6**(11): a019331.

Guenther, M. G., S. S. Levine, L. A. Boyer, R. Jaenisch and R. A. Young (2007). "A chromatin landmark and transcription initiation at most promoters in human cells." Cell **130**(1): 77-88.

Hah, N., S. Murakami, A. Nagari, C. G. Danko and W. L. Kraus (2013). "Enhancer transcripts mark active estrogen receptor binding sites." Genome Res **23**(8): 1210-1223.

Hampsey, M. (1998). "Molecular genetics of the RNA polymerase II general transcriptional machinery." Microbiol Mol Biol Rev **62**(2): 465-503.

Harrow, J., A. Frankish, J. M. Gonzalez, E. Tapanari, M. Diekhans, F. Kokocinski, B. L. Aken, D. Barrell, A. Zadissa, S. Searle, I. Barnes, A. Bignell, V. Boychenko, T. Hunt, M. Kay, G. Mukherjee, J. Rajan, G. Despacio-Reyes, G. Saunders, C. Steward, R. Harte, M. Lin, C. Howald, A. Tanzer, T. Derrien, J. Chrast, N. Walters, S. Balasubramanian, B. Pei, M. Tress, J. M. Rodriguez, I. Ezkurdia, J. van Baren, M. Brent, D. Haussler, M. Kellis, A. Valencia, A. Reymond, M. Gerstein, R. Guigo and T. J. Hubbard (2012). "GENCODE: the reference human genome annotation for The ENCODE Project." Genome Res **22**(9): 1760-1774.

Hawkins, R. D., G. C. Hon, L. K. Lee, Q. Ngo, R. Lister, M. Pelizzola, L. E. Edsall, S. Kuan, Y. Luu, S. Klugman, J. Antosiewicz-Bourget, Z. Ye, C. Espinoza, S. Agarwahl, L. Shen, V. Ruotti, W. Wang, R. Stewart, J. A. Thomson, J. R. Ecker and B. Ren (2010). "Distinct epigenomic landscapes of pluripotent and lineage-committed human cells." Cell Stem Cell **6**(5): 479-491.

Heintzman, N. D., G. C. Hon, R. D. Hawkins, P. Kheradpour, A. Stark, L. F. Harp, Z. Ye, L. K. Lee, R. K. Stuart, C. W. Ching, K. A. Ching, J. E. Antosiewicz-Bourget, H. Liu, X. Zhang, R. D. Green, V. V. Lobanenkov, R. Stewart, J. A. Thomson, G. E. Crawford, M. Kellis and B. Ren (2009). "Histone modifications at human enhancers reflect global cell-type-specific gene expression." Nature **459**(7243): 108-112.

Heintzman, N. D., R. K. Stuart, G. Hon, Y. Fu, C. W. Ching, R. D. Hawkins, L. O. Barrera, S. Van Calcar, C. Qu, K. A. Ching, W. Wang, Z. Weng, R. D. Green, G. E. Crawford and B. Ren (2007). "Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome." Nat Genet **39**(3): 311-318.

Heinz, S., C. Benner, N. Spann, E. Bertolino, Y. C. Lin, P. Laslo, J. X. Cheng, C. Murre, H. Singh and C. K. Glass (2010). "Simple Combinations of Lineage-Determining Transcription Factors Prime cis-Regulatory Elements Required for Macrophage and B Cell Identities." Molecular Cell **38**(4): 576-589.

Heinz, S., C. E. Romanoski, C. Benner and C. K. Glass (2015). "The selection and function of cell type-specific enhancers." Nat Rev Mol Cell Biol **16**(3): 144-154.

Henikoff, S., J. G. Henikoff, A. Sakai, G. B. Loeb and K. Ahmad (2009). "Genome-wide profiling of salt fractions maps physical properties of chromatin." Genome Res **19**(3): 460-469.

Hetzel, J., S. H. Duttke, C. Benner and J. Chory (2016). "Nascent RNA sequencing reveals distinct features in plant transcription." Proc Natl Acad Sci U S A **113**(43): 12316-12321.

Hirano, T. (2012). "Condensins: universal organizers of chromosomes with diverse functions." Genes Dev **26**(15): 1659-1678.

Hnisz, D., K. Shrinivas, R. A. Young, A. K. Chakraborty and P. A. Sharp (2017). "A Phase Separation Model for Transcriptional Control." Cell **169**(1): 13-23.

Hodl, M. and K. Basler (2009). "Transcription in the absence of histone H3.3." Curr Biol **19**(14): 1221-1226.

Hodl, M. and K. Basler (2012). "Transcription in the absence of histone H3.2 and H3K4 methylation." Curr Biol **22**(23): 2253-2257.

Hsieh, C. L., T. Fei, Y. Chen, T. Li, Y. Gao, X. Wang, T. Sun, C. J. Sweeney, G. S. Lee, S. Chen, S. P. Balk, X. S. Liu, M. Brown and P. W. Kantoff (2014). "Enhancer RNAs participate in androgen receptor-driven looping that selectively enhances gene activation." Proc Natl Acad Sci U S A **111**(20): 7319-7324.

Jiang, C. and B. F. Pugh (2009). "Nucleosome positioning and gene regulation: advances through genomics." Nat Rev Genet **10**(3): 161-172.

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

Jin, C. and G. Felsenfeld (2007). "Nucleosome stability mediated by histone variants H3.3 and H2A.Z." Genes Dev **21**(12): 1519-1529.

Jin, C., C. Zang, G. Wei, K. Cui, W. Peng, K. Zhao and G. Felsenfeld (2009). "H3.3/H2A.Z double variant-containing nucleosomes mark 'nucleosome-free regions' of active promoters and other regulatory regions." Nat Genet **41**(8): 941-945.

Jin, Y., U. Eser, K. Struhl and L. S. Churchman (2017). "The Ground State and Evolution of Promoter Region Directionality." Cell **170**(5): 889-898.e810.

Kadonaga, J. T. (2012). "Perspectives on the RNA polymerase II core promoter." Wiley Interdiscip Rev Dev Biol **1**(1): 40-51.

Kagey, M. H., J. J. Newman, S. Bilodeau, Y. Zhan, D. A. Orlando, N. L. van Berkum, C. C. Ebmeier, J. Goossens, P. B. Rahl, S. S. Levine, D. J. Taatjes, J. Dekker and R. A. Young (2010). "Mediator and cohesin connect gene expression and chromatin architecture." Nature **467**(7314): 430-435.

Kaikkonen, M. U., N. J. Spann, S. Heinz, C. E. Romanoski, K. A. Allison, J. D. Stender, H. B. Chun, D. F. Tough, R. K. Prinjha, C. Benner and C. K. Glass (2013). "Remodeling of the enhancer landscape during macrophage activation is coupled to enhancer transcription." Mol Cell **51**(3): 310-325.

Kapranov, P., J. Cheng, S. Dike, D. A. Nix, R. Duttagupta, A. T. Willingham, P. F. Stadler, J. Hertel, J. Hackermuller, I. L. Hofacker, I. Bell, E. Cheung, J. Drenkow, E. Dumais, S. Patel, G. Helt, M. Ganesh, S. Ghosh, A. Piccolboni, V. Sementchenko, H. Tammana and T. R. Gingeras (2007). "RNA maps reveal new RNA classes and a possible function for pervasive transcription." Science **316**(5830): 1484-1488.

Keene, M. A., V. Corces, K. Lowenhaupt and S. C. Elgin (1981). "DNase I hypersensitive sites in Drosophila chromatin occur at the 5' ends of regions of transcription." Proc Natl Acad Sci U S A **78**(1): 143-146.

Khoury, G. and P. Gruss (1983). "Enhancer elements." Cell **33**(2): 313-314.

Kim, T. H., L. O. Barrera, M. Zheng, C. Qu, M. A. Singer, T. A. Richmond, Y. Wu, R. D. Green and B. Ren (2005). "A high-resolution map of active promoters in the human genome." Nature **436**(7052): 876-880.

Kim, T. K., M. Hemberg, J. M. Gray, A. M. Costa, D. M. Bear, J. Wu, D. A. Harmin, M. Laptewicz, K. Barbara-Haley, S. Kuersten, E. Markenscoff-Papadimitriou, D. Kuhl, H. Bito, P. F. Worley, G. Kreiman and M. E. Greenberg (2010). "Widespread transcription at neuronal activity-regulated enhancers." Nature **465**(7295): 182-187.

Kim, T. K. and R. Shiekhattar (2015). "Architectural and Functional Commonalities between Enhancers and Promoters." Cell **162**(5): 948-959.

Klein, J. C., A. Keith, V. Agarwal, T. Durham and J. Shendure (2018). "Functional characterization of enhancer evolution in the primate lineage." Genome Biol **19**(1): 99.

Kleinjan, D. A. and V. van Heyningen (2005). "Long-range control of gene expression: emerging mechanisms and disruption in disease." Am J Hum Genet **76**(1): 8-32.

Koch, F., R. Fenouil, M. Gut, P. Cauchy, T. K. Albert, J. Zacarias-Cabeza, S. Spicuglia, A. L. de la Chapelle, M. Heidemann, C. Hintermair, D. Eick, I. Gut, P. Ferrier and J. C. Andrau (2011). "Transcription initiation platforms and GTF recruitment at tissue-specific enhancers and promoters." Nat Struct Mol Biol **18**(8): 956-963.

Kornberg, R. D. (1977). "Structure of chromatin." Annu Rev Biochem **46**: 931-954.

Kowalczyk, M. S., J. R. Hughes, D. Garrick, M. D. Lynch, J. A. Sharpe, J. A. Sloane-Stanley, S. J. McGowan, M. De Gobbi, M. Hosseini, D. Vernimmen, J. M. Brown, N. E. Gray, L. Collavin, R. J. Gibbons, J. Flint, S. Taylor, V. J. Buckle, T. A. Milne, W. G. Wood and D. R. Higgs (2012). "Intragenic enhancers act as alternative promoters." Mol Cell **45**(4): 447-458.

Kulaeva, O. I., E. V. Nizovtseva, Y. S. Polikanov, S. V. Ulianov and V. M. Studitsky (2012). "Distant activation of transcription: mechanisms of enhancer action." Mol Cell Biol **32**(24): 4892-4897.

Kvon, E. Z., T. Kazmar, G. Stampfel, J. O. Yanez-Cuna, M. Pagani, K. Schernhuber, B. J. Dickson and A. Stark (2014). "Genome-scale functional characterization of Drosophila developmental enhancers in vivo." Nature **512**(7512): 91-95.

Lai, F. and R. Shiekhattar (2014). "Enhancer RNAs: the new molecules of transcription." Curr Opin Genet Dev **25**: 38-42.

Lai, W. K. and B. F. Pugh (2017). "Genome-wide uniformity of human 'open' pre-initiation complexes." Genome Res **27**(1): 15-26.

Lee, L. R., D. L. Wengier and D. C. Bergmann (2019). "Cell-type-specific transcriptome and histone modification dynamics during cellular reprogramming in the Arabidopsis stomatal lineage." Proc Natl Acad Sci U S A **116**(43): 21914-21924.

Lee, T. I. and R. A. Young (2000). "Transcription of eukaryotic protein-coding genes." Annu Rev Genet **34**: 77-137.

Lemon, B. and R. Tjian (2000). "Orchestrated response: a symphony of transcription factors for gene control." Genes Dev **14**(20): 2551-2569.

Lettice, L. A., S. J. Heaney, L. A. Purdie, L. Li, P. de Beer, B. A. Oostra, D. Goode, G. Elgar, R. E. Hill and E. de Graaff (2003). "A long-range Shh enhancer regulates expression in the developing limb and fin and is associated with preaxial polydactyly." Hum Mol Genet **12**(14): 1725-1735.

Li, B., M. Carey and J. L. Workman (2007). "The role of chromatin during transcription." Cell **128**(4): 707-719.

Li, G., X. Ruan, R. K. Auerbach, K. S. Sandhu, M. Zheng, P. Wang, H. M. Poh, Y. Goh, J. Lim, J. Zhang, H. S. Sim, S. Q. Peh, F. H. Mulawadi, C. T. Ong, Y. L. Orlov, S. Hong, Z. Zhang, S. Landt, D. Raha, G. Euskirchen, C. L. Wei, W. Ge, H. Wang, C. Davis, K. I. Fisher-Aylor, A. Mortazavi, M. Gerstein, T. Gingeras, B. Wold, Y. Sun, M. J. Fullwood, E. Cheung, E. Liu, W. K. Sung, M. Snyder and Y. Ruan (2012). "Extensive promoter-centered chromatin interactions provide a topological basis for transcription regulation." Cell **148**(1-2): 84-98.

Li, S., M. Yamada, X. Han, U. Ohler and P. N. Benfey (2016). "High-Resolution Expression Map of the Arabidopsis Root Reveals Alternative Splicing and lincRNA Regulation." Developmental Cell **39**(4): 508-522.

Li, W., D. Notani, Q. Ma, B. Tanasa, E. Nunez, A. Y. Chen, D. Merkurjev, J. Zhang, K. Ohgi, X. Song, S. Oh, H. S. Kim, C. K. Glass and M. G. Rosenfeld (2013). "Functional roles of enhancer RNAs for oestrogen-dependent transcriptional activation." Nature **498**(7455): 516-520.

Lifton, R. P., M. L. Goldberg, R. W. Karp and D. S. Hogness (1978). "The organization of the histone genes in Drosophila melanogaster: functional and evolutionary implications." Cold Spring Harb Symp Quant Biol **42 Pt 2**: 1047-1051.

Liu, C., C. Wang, G. Wang, C. Becker, M. Zaidem and D. Weigel (2016). "Genome-wide analysis of chromatin packing in Arabidopsis thaliana at single-gene resolution." Genome Res **26**(8): 1057-1068.

Liu, J., C. Jung, J. Xu, H. Wang, S. Deng, L. Bernad, C. Arenas-Huertero and N. H. Chua (2012). "Genome-wide analysis uncovers regulation of long intergenic noncoding RNAs in Arabidopsis." Plant Cell **24**(11): 4333-4345.

Liu, S., Y. Liu, Q. Zhang, J. Wu, J. Liang, S. Yu, G. H. Wei, K. P. White and X. Wang (2017). "Systematic identification of regulatory variants associated with cancer risk." Genome Biol **18**(1): 194.

Liu, Y., S. Yu, V. K. Dhiman, T. Brunetti, H. Eckart and K. P. White (2017). "Functional assessment of human enhancer activities using whole-genome STARR-sequencing." Genome Biol **18**(1): 219.

Lowndes, N. F. and G. W. Toh (2005). "DNA repair: the importance of phosphorylating histone H2AX." Curr Biol **15**(3): R99-r102.

Lu, Z., A. P. Marand, W. A. Ricci, C. L. Ethridge, X. Zhang and R. J. Schmitz (2019). "The prevalence, evolution and chromatin signatures of plant regulatory elements." Nat Plants **5**(12): 1250-1259.

Luger, K., M. L. Dechassa and D. J. Tremethick (2012). "New insights into nucleosome and chromatin structure: an ordered state or a disordered affair?" Nat Rev Mol Cell Biol **13**(7): 436-447.

Luger, K., A. Mader, D. F. Sargent and T. J. Richmond (2000). "The atomic structure of the nucleosome core particle." J Biomol Struct Dyn **17 Suppl 1**: 185-188.

Lyons, E. and M. Freeling (2008). "How to usefully compare homologous plant genes and chromosomes as DNA sequences." Plant J **53**(4): 661-673.

Lyons, E., B. Pedersen, J. Kane and M. Freeling (2008). "The Value of Nonmodel Genomes and an Example Using SynMap Within CoGe to Dissect the Hexaploidy that Predates the Rosids." Tropical Plant Biology **1**(3): 181-190.

Maher, K. A., M. Bajic, K. Kajala, M. Reynoso, G. Pauluzzi, D. A. West, K. Zumstein, M. Woodhouse, K. Bubb, M. W. Dorrity, C. Queitsch, J. Bailey-Serres, N. Sinha, S. M. Brady and R. B. Deal (2018). "Profiling of Accessible Chromatin Regions across Multiple Plant Species and Cell Types Reveals Common Gene Regulatory Principles and New Control Modules." Plant Cell **30**(1): 15-36.

Martienssen, R. and D. Moazed (2015). "RNAi and heterochromatin assembly." Cold Spring Harb Perspect Biol **7**(8): a019323.

Marzluff, W. F. and R. J. Duronio (2002). "Histone mRNA expression: multiple levels of cell cycle regulation and important developmental consequences." Curr Opin Cell Biol **14**(6): 692-699.

Marzluff, W. F., P. Gongidi, K. R. Woods, J. Jin and L. J. Maltais (2002). "The human and mouse replication-dependent histone genes." Genomics **80**(5): 487-498.

Mathelier, A., X. Zhao, A. W. Zhang, F. Parcy, R. Worsley-Hunt, D. J. Arenillas, S. Buchman, C. Y. Chen, A. Chou, H. Ienasescu, J. Lim, C. Shyr, G. Tan, M. Zhou, B. Lenhard, A. Sandelin and W. W. Wasserman (2014). "JASPAR 2014: an extensively expanded and updated open-access database of transcription factor binding profiles." Nucleic Acids Res **42**(Database issue): D142-147.

Mathur, J. and C. Koncz (1998). "PEG-mediated protoplast transformation with naked DNA." Methods Mol Biol **82**: 267-276.

Mavrich, T. N., C. Jiang, I. P. Ioshikhes, X. Li, B. J. Venters, S. J. Zanton, L. P. Tomsho, J. Qi, R. L. Glaser, S. C. Schuster, D. S. Gilmour, I. Albert and B. F. Pugh (2008). "Nucleosome organization in the Drosophila genome." Nature **453**(7193): 358-362.

McGarry, R. C. and B. G. Ayre (2008). "A DNA element between At4g28630 and At4g28640 confers companion-cell specific expression following the sink-to-source transition in mature minor vein phloem." Planta **228**(5): 839-849.

McGhee, J. D., W. I. Wood, M. Dolan, J. D. Engel and G. Felsenfeld (1981). "A 200 base pair region at the 5' end of the chicken adult beta-globin gene is accessible to nuclease digestion." Cell **27**(1 Pt 2): 45-55.

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

Mendenhall, E. M., K. E. Williamson, D. Reyon, J. Y. Zou, O. Ram, J. K. Joung and B. E. Bernstein (2013). "Locus-specific editing of histone modifications at endogenous enhancers." Nat Biotechnol **31**(12): 1133-1136.

Meyer, P. (2011). "DNA methylation systems and targets in plants." FEBS Lett **585**(13): 2008-2015.

Mieczkowski, J., A. Cook, S. K. Bowman, B. Mueller, B. H. Alver, S. Kundu, A. M. Deaton, J. A. Urban, E. Larschan, P. J. Park, R. E. Kingston and M. Y. Tolstorukov (2016). "MNase titration reveals differences between nucleosome occupancy and chromatin accessibility." Nat Commun **7**: 11485.

Mikhaylichenko, O., V. Bondarenko, D. Harnett, I. E. Schor, M. Males, R. R. Viales and E. E. M. Furlong (2018). "The degree of enhancer or promoter activity is reflected by the levels and directionality of eRNA transcription." Genes Dev **32**(1): 42-57.

Moreau, P., R. Hen, B. Wasylyk, R. Everett, M. P. Gaub and P. Chambon (1981). "The SV40 72 base repair repeat has a striking effect on gene expression both in SV40 and other chimeric recombinants." Nucleic Acids Res **9**(22): 6047-6068.

Morrison, A. J. and X. Shen (2009). "Chromatin remodelling beyond transcription: the INO80 and SWR1 complexes." Nat Rev Mol Cell Biol **10**(6): 373-384.

Mousavi, K., H. Zare, S. Dell'orso, L. Grontved, G. Gutierrez-Cruz, A. Derfoul, G. L. Hager and V. Sartorelli (2013). "eRNAs promote transcription by establishing chromatin accessibility at defined genomic loci." Mol Cell **51**(5): 606-617.

Mueller, B., J. Mieczkowski, S. Kundu, P. Wang, R. Sadreyev, M. Y. Tolstorukov and R. E. Kingston (2017). "Widespread changes in nucleosome accessibility without changes in nucleosome occupancy during a rapid transcriptional induction." Genes Dev **31**(5): 451-462.

Muerdter, F., L. M. Boryn and C. D. Arnold (2015). "STARR-seq - principles and applications." Genomics **106**(3): 145-150.

Muerdter, F., L. M. Boryn, A. R. Woodfin, C. Neumayr, M. Rath, M. A. Zabidi, M. Pagani, V. Haberle, T. Kazmar, R. R. Catarino, K. Schernhuber, C. D. Arnold and A. Stark (2018). "Resolving systematic errors in widely used enhancer activity assays in human cells." Nat Methods **15**(2): 141-149.

Muerdter, F. and A. Stark (2016). "Gene Regulation: Activation through Space." Curr Biol **26**(19): R895-r898.

Mumbach, M. R., A. T. Satpathy, E. A. Boyle, C. Dai, B. G. Gowen, S. W. Cho, M. L. Nguyen, A. J. Rubin, J. M. Granja, K. R. Kazane, Y. Wei, T. Nguyen, P. G. Greenside, M. R. Corces, J. Tycko, D. R. Simeonov, N. Suliman, R. Li, J. Xu, R. A. Flynn, A. Kundaje, P. A. Khavari, A. Marson, J. E. Corn, T. Quertermous, W. J. Greenleaf and H. Y. Chang (2017). "Enhancer connectome in primary human cells identifies target genes of disease-associated DNA elements." Nat Genet **49**(11): 1602-1612.

Nasmyth, K. and C. H. Haering (2009). "Cohesin: its roles and mechanisms." Annu Rev Genet **43**: 525-558.

Natoli, G. and J. C. Andrau (2012). "Noncoding transcription at enhancers: general principles and functional models." Annu Rev Genet **46**: 1-19.

Neuberger, M. S. (1983). "Expression and regulation of immunoglobulin heavy chain gene transfected into lymphoid cells." Embo j **2**(8): 1373-1378.

Nguyen, T. A., R. D. Jones, A. R. Snavely, A. R. Pfenning, R. Kirchner, M. Hemberg and J. M. Gray (2016). "High-throughput functional comparison of promoter and enhancer activities." Genome Res **26**(8): 1023-1033.

O'Malley, R. C., S. C. Huang, L. Song, M. G. Lewsey, A. Bartlett, J. R. Nery, M. Galli, A. Gallavotti and J. R. Ecker (2016). "Cistrome and Epicistrome Features Shape the Regulatory DNA Landscape." Cell **165**(5): 1280-1292.

O'Malley, R. C., S. C. Huang, L. Song, M. G. Lewsey, A. Bartlett, J. R. Nery, M. Galli, A. Gallavotti and J. R. Ecker (2016). "Cistrome and Epicistrome Features Shape the Regulatory DNA Landscape." Cell **166**(6): 1598.

Oka, R., J. Zicola, B. Weber, S. N. Anderson, C. Hodgman, J. I. Gent, J. J. Wesselink, N. M. Springer, H. C. J. Hoefsloot, F. Turck and M. Stam (2017). "Genome-wide mapping of transcriptional enhancer candidates using DNA and chromatin features in maize." Genome Biol **18**(1): 137.

Olins, A. L. and D. E. Olins (1974). "Spheroid chromatin units (v bodies)." Science **183**(4122): 330-332.

Ong, C. T. and V. G. Corces (2011). "Enhancer function: new insights into the regulation of tissue-specific gene expression." Nat Rev Genet **12**(4): 283-293.

Oyama, T., Y. Shimura and K. Okada (1997). "The Arabidopsis HY5 gene encodes a bZIP protein that regulates stimulus-induced development of root and hypocotyl." Genes Dev **11**(22): 2983-2995.

Pajoro, A., P. Madrigal, J. M. Muino, J. T. Matus, J. Jin, M. A. Mecchia, J. M. Debernardi, J. F. Palatnik, S. Balazadeh, M. Arif, D. S. O'Maoileidigh, F. Wellmer, P. Krajewski, J. L. Riechmann, G. C. Angenent and K. Kaufmann (2014). "Dynamics of chromatin accessibility and gene regulation by MADS-domain transcription factors in flower development." Genome Biol **15**(3): R41.

Palmer, D. K., K. O'Day, H. L. Trong, H. Charbonneau and R. L. Margolis (1991). "Purification of the centromere-specific protein CENP-A and demonstration that it is a distinctive histone." Proc Natl Acad Sci U S A **88**(9): 3734-3738.

Pekowska, A., T. Benoukraf, J. Zacarias-Cabeza, M. Belhocine, F. Koch, H. Holota, J. Imbert, J. C. Andrau, P. Ferrier and S. Spicuglia (2011). "H3K4 tri-methylation provides an epigenetic signature of active enhancers." Embo j **30**(20): 4198-4210.

Petrykowska, H. M., C. M. Vockley and L. Elnitski (2008). "Detection and characterization of silencers and enhancer-blockers in the greater CFTR locus." Genome Res **18**(8): 1238-1246.

Pikaard, C. S. and O. Mittelsten Scheid (2014). "Epigenetic regulation in plants." Cold Spring Harb Perspect Biol **6**(12): a019315.

Pollex, T. and E. E. M. Furlong (2017). "Correlation Does Not Imply Causation: Histone Methyltransferases, but Not Histone Methylation, SET the Stage for Enhancer Activation." Mol Cell **66**(4): 439-441.

Poss, Z. C., C. C. Ebmeier and D. J. Taatjes (2013). "The Mediator complex and transcription regulation." Crit Rev Biochem Mol Biol **48**(6): 575-608.

Preker, P., J. Nielsen, S. Kammler, S. Lykke-Andersen, M. S. Christensen, C. K. Mapendano, M. H. Schierup and T. H. Jensen (2008). "RNA exosome depletion reveals transcription upstream of active human promoters." Science **322**(5909): 1851-1854.

Rada-Iglesias, A., R. Bajpai, T. Swigut, S. A. Brugmann, R. A. Flynn and J. Wysocka (2011). "A unique chromatin signature uncovers early developmental enhancers in humans." Nature **470**(7333): 279-283.

Ricci, W. A., Z. Lu, L. Ji, A. P. Marand, C. L. Ethridge, N. G. Murphy, J. M. Noshay, M. Galli, M. K. Mejia-Guerra, M. Colome-Tatche, F. Johannes, M. J. Rowley, V. G. Corces, J. Zhai, M. J. Scanlon, E. S. Buckler, A. Gallavotti, N. M. Springer, R. J. Schmitz and X. Zhang (2019). "Widespread long-range cis-regulatory elements in the maize genome." Nat Plants **5**(12): 1237-1249.

Rickels, R., H. M. Herz, C. C. Sze, K. Cao, M. A. Morgan, C. K. Collings, M. Gause, Y. H. Takahashi, L. Wang, E. J. Rendleman, S. A. Marshall, A. Krueger, E. T. Bartom, A. Piunti, E. R. Smith, N. A. Abshiru, N. L. Kelleher, D. Dorsett and A. Shilatifard (2017). "Histone H3K4 monomethylation catalyzed by Trr and mammalian COMPASS-like proteins at enhancers is dispensable for development and viability." Nat Genet **49**(11): 1647-1653.

Robinson, P. J. and D. Rhodes (2006). "Structure of the '30 nm' chromatin fibre: a key role for the linker histone." Curr Opin Struct Biol **16**(3): 336-343.

Rodgers-Melnick, E., D. L. Vera, H. W. Bass and E. S. Buckler (2016). "Open chromatin reveals the functional maize genome." Proc Natl Acad Sci U S A **113**(22): E3177-3184.

Roh, T. Y., S. Cuddapah and K. Zhao (2005). "Active chromatin domains are defined by acetylation islands revealed by genome-wide mapping." Genes Dev **19**(5): 542-552.

Roh, T. Y., G. Wei, C. M. Farrell and K. Zhao (2007). "Genome-wide prediction of conserved and nonconserved enhancers by histone acetylation patterns." Genome Res **17**(1): 74-81.

Sainsbury, S., C. Bernecky and P. Cramer (2015). "Structural basis of transcription initiation by RNA polymerase II." Nat Rev Mol Cell Biol **16**(3): 129-143.

Saxonov, S., P. Berg and D. L. Brutlag (2006). "A genome-wide analysis of CpG dinucleotides in the human genome distinguishes two distinct classes of promoters." Proc Natl Acad Sci U S A **103**(5): 1412-1417.

Schaukowitch, K., J. Y. Joo, X. Liu, J. K. Watts, C. Martinez and T. K. Kim (2014). "Enhancer RNA facilitates NELF release from immediate early genes." Mol Cell **56**(1): 29-42.

Schmidt, D., P. C. Schwalie, C. S. Ross-Innes, A. Hurtado, G. D. Brown, J. S. Carroll, P. Flicek and D. T. Odom (2010). "A CTCF-independent role for cohesin in tissue-specific transcription." Genome Res **20**(5): 578-588.

Schone, S., M. Bothe, E. Einfeldt, M. Borschiwer, P. Benner, M. Vingron, M. Thomas-Chollier and S. H. Meijsing (2018). "Synthetic STARR-seq reveals how DNA shape and sequence modulate transcriptional output and noise." PLoS Genet **14**(11): e1007793.

Schones, D. E., K. Cui, S. Cuddapah, T. Y. Roh, A. Barski, Z. Wang, G. Wei and K. Zhao (2008). "Dynamic regulation of nucleosome positioning in the human genome." Cell **132**(5): 887-898.

Schwaiger, M., A. Schonauer, A. F. Rendeiro, C. Pribitzer, A. Schauer, A. F. Gilles, J. B. Schinko, E. Renfer, D. Fredman and U. Technau (2014). "Evolutionary conservation of the eumetazoan gene regulatory landscape." Genome Res **24**(4): 639-650.

Scruggs, B. S., D. A. Gilchrist, S. Nechaev, G. W. Muse, A. Burkholder, D. C. Fargo and K. Adelman (2015). "Bidirectional Transcription Arises from Two Distinct Hubs of Transcription Factor Binding and Active Chromatin." Mol Cell **58**(6): 1101-1112.

Seila, A. C., J. M. Calabrese, S. S. Levine, G. W. Yeo, P. B. Rahl, R. A. Flynn, R. A. Young and P. A. Sharp (2008). "Divergent transcription from active promoters." Science **322**(5909): 1849-1851.

Seila, A. C., L. J. Core, J. T. Lis and P. A. Sharp (2009). "Divergent transcription: a new feature of active promoters." Cell Cycle **8**(16): 2557-2564.

Serfling, E., A. Lubbe, K. Dorsch-Hasler and W. Schaffner (1985). "Metal-dependent SV40 viruses containing inducible enhancers from the upstream region of metallothionein genes." Embo j **4**(13b): 3851-3859.

Sherwood, R. I., T. Hashimoto, C. W. O'Donnell, S. Lewis, A. A. Barkal, J. P. van Hoff, V. Karun, T. Jaakkola and D. K. Gifford (2014). "Discovery of directional and nondirectional pioneer transcription factors by modeling DNase profile magnitude and shape." Nat Biotechnol **32**(2): 171-178.

Shilatifard, A. (2006). "Chromatin modifications by methylation and ubiquitination: implications in the regulation of gene expression." Annu Rev Biochem **75**: 243-269.

Shlyueva, D., G. Stampfel and A. Stark (2014). "Transcriptional enhancers: from properties to genome-wide predictions." Nat Rev Genet **15**(4): 272-286.

Sijacic, P., M. Bajic, E. C. McKinney, R. B. Meagher and R. B. Deal (2018). "Changes in chromatin accessibility between Arabidopsis stem cells and mesophyll cells illuminate cell type-specific transcription factor networks." Plant J **94**(2): 215-231.

Simpson, J., V. A. N. M. M and L. Herrera-Estrella (1986). "Photosynthesis-associated gene families: differences in response to tissue-specific and environmental factors." Science **233**(4759): 34-38.

Sims, R. J., 3rd, R. Belotserkovskaya and D. Reinberg (2004). "Elongation by RNA polymerase II: the short and long of it." Genes Dev **18**(20): 2437-2468.

Smale, S. T. and D. Baltimore (1989). "The "initiator" as a transcription control element." Cell **57**(1): 103-113.

Song, F., P. Chen, D. Sun, M. Wang, L. Dong, D. Liang, R. M. Xu, P. Zhu and G. Li (2014). "Cryo-EM study of the chromatin fiber reveals a double helix twisted by tetranucleosomal units." Science **344**(6182): 376-380.

Soufi, A., M. F. Garcia, A. Jaroszewicz, N. Osman, M. Pellegrini and K. S. Zaret (2015). "Pioneer transcription factors target partial DNA motifs on nucleosomes to initiate reprogramming." Cell **161**(3): 555-568.

Spitz, F. and E. E. Furlong (2012). "Transcription factors: from enhancer binding to developmental control." Nat Rev Genet **13**(9): 613-626.

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

Strome, S., W. G. Kelly, S. Ercan and J. D. Lieb (2014). "Regulation of the X chromosomes in Caenorhabditis elegans." Cold Spring Harb Perspect Biol **6**(3).

Struhl, K. (2007). "Transcriptional noise and the fidelity of initiation by RNA polymerase II." Nat Struct Mol Biol **14**(2): 103-105.

Studer, A., Q. Zhao, J. Ross-Ibarra and J. Doebley (2011). "Identification of a functional transposon insertion in the maize domestication gene tb1." Nat Genet **43**(11): 1160-1163.

Sullivan, A. M., A. A. Arsovski, J. Lempe, K. L. Bubb, M. T. Weirauch, P. J. Sabo, R. Sandstrom, R. E. Thurman, S. Neph, A. P. Reynolds, A. B. Stergachis, B. Vernot, A. K. Johnson, E. Haugen, S. T. Sullivan, A. Thompson, F. V. Neri, 3rd, M. Weaver, M. Diegel, S. Mnaimneh, A. Yang, T. R. Hughes, J. L. Nemhauser, C. Queitsch and J. A. Stamatoyannopoulos (2014). "Mapping and dynamics of regulatory DNA and transcription factor networks in A. thaliana." Cell Rep **8**(6): 2015-2030.

Tan-Wong, S. M., J. B. Zaugg, J. Camblong, Z. Xu, D. W. Zhang, H. E. Mischo, A. Z. Ansari, N. M. Luscombe, L. M. Steinmetz and N. J. Proudfoot (2012). "Gene loops enhance transcriptional directionality." Science **338**(6107): 671-675.

Timko, M. P., A. P. Kausch, C. Castresana, J. Fassler, L. Herrera-Estrella, G. Van den Broeck, M. Van Montagu, J. Schell and A. R. Cashmore (1985). "Light regulation of plant gene expression by an upstream enhancer-like element." Nature **318**(6046): 579-582.

Tolhuis, B., R. J. Palstra, E. Splinter, F. Grosveld and W. de Laat (2002). "Looping and interaction between hypersensitive sites in the active beta-globin locus." Mol Cell **10**(6): 1453-1465.

Trinklein, N. D., S. F. Aldred, S. J. Hartman, D. I. Schroeder, R. P. Otillar and R. M. Myers (2004). "An abundance of bidirectional promoters in the human genome." Genome Res **14**(1): 62-66.

Tsai, A., A. K. Muthusamy, M. R. Alves, L. D. Lavis, R. H. Singer, D. L. Stern and J. Crocker (2017). "Nuclear microenvironments modulate transcription from low-affinity enhancers." Elife **6**.

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

Valles, M. B., J; Azorin, F; Puigdomenech, P (1991). "Nuclease sensitivity of a maize HRGP gene in chromatin and in naked DNA." Plant Science **78**(2): 225-230.

van Arensbergen, J., V. D. FitzPatrick, M. de Haas, L. Pagie, J. Sluimer, H. J. Bussemaker and B. van Steensel (2017). "Genome-wide mapping of autonomous promoter activity in human cells." Nat Biotechnol **35**(2): 145-153.

Vernimmen, D. and W. A. Bickmore (2015). "The Hierarchy of Transcriptional Activation: From Enhancer to Promoter." Trends Genet **31**(12): 696-708.

Villar, D., C. Berthelot, S. Aldridge, T. F. Rayner, M. Lukk, M. Pignatelli, T. J. Park, R. Deaville, J. T. Erichsen, A. J. Jasinska, J. M. Turner, M. F. Bertelsen, E. P. Murchison, P. Flicek and D. T. Odom (2015). "Enhancer evolution across 20 mammalian species." Cell **160**(3): 554-566.

Vokes, S. A., H. Ji, W. H. Wong and A. P. McMahon (2008). "A genome-scale analysis of the cis-regulatory circuitry underlying sonic hedgehog-mediated patterning of the mammalian limb." Genes Dev **22**(19): 2651-2663.

Voss, T. C. and G. L. Hager (2014). "Dynamic regulation of transcriptional states by chromatin and transcription factors." Nat Rev Genet **15**(2): 69-81.

Wang, C., C. Liu, D. Roqueiro, D. Grimm, R. Schwab, C. Becker, C. Lanz and D. Weigel (2015). "Genome-wide analysis of local chromatin packing in Arabidopsis thaliana." Genome Res **25**(2): 246-256.

Wang, D. and R. B. Deal (2015). "Epigenome profiling of specific plant cell types using a streamlined INTACT protocol and ChIP-seq." Methods Mol Biol **1284**: 3-25.

Wang, H., P. J. Chung, J. Liu, I. C. Jang, M. J. Kean, J. Xu and N. H. Chua (2014). "Genome-wide identification of long noncoding natural antisense transcripts and their responses to light in Arabidopsis." Genome Res **24**(3): 444-453.

Wang, M., L. Tu, M. Lin, Z. Lin, P. Wang, Q. Yang, Z. Ye, C. Shen, J. Li, L. Zhang, X. Zhou, X. Nie, Z. Li, K. Guo, Y. Ma, C. Huang, S. Jin, L. Zhu, X. Yang, L. Min, D. Yuan, Q. Zhang, K. Lindsey and X. Zhang (2017). "Asymmetric subgenome selection and cis-regulatory divergence during cotton domestication." Nat Genet **49**(4): 579-587.

Wang, X., L. He, S. M. Goggin, A. Saadat, L. Wang, N. Sinnott-Armstrong, M. Claussnitzer and M. Kellis (2018). "High-resolution genome-wide functional dissection of transcriptional regulatory regions and nucleotides in human." Nat Commun **9**(1): 5380.

Wang, Z., C. Zang, J. A. Rosenfeld, D. E. Schones, A. Barski, S. Cuddapah, K. Cui, T. Y. Roh, W. Peng, M. Q. Zhang and K. Zhao (2008). "Combinatorial patterns of histone acetylations and methylations in the human genome." Nat Genet **40**(7): 897-903.

Weber, B., J. Zicola, R. Oka and M. Stam (2016). "Plant Enhancers: A Call for Discovery." Trends Plant Sci **21**(11): 974-987.

Weirauch, M. T., A. Yang, M. Albu, A. Cote, A. Montenegro-Montero, P. Drewe, H. S. Najafabadi, S. A. Lambert, I. Mann, K. Cook, H. Zheng, A. Goity, H. van Bakel, J. C. Lozano, M. Galli, M. Lewsey, E. Huang, T. Mukherjee, X. Chen, J. S. Reece-Hoyes, S. Govindarajan, G. Shaulsky, A. J. M. Walhout, F. Y. Bouget, G. Ratsch, L. F. Larrondo, J. R. Ecker and T. R. Hughes (2014). "Determination and inference of eukaryotic transcription factor sequence specificity." Cell **158**(6): 1431-1443.

Weirauch, M. T., A. Yang, M. Albu, A. G. Cote, A. Montenegro-Montero, P. Drewe, H. S. Najafabadi, S. A. Lambert, I. Mann, K. Cook, H. Zheng, A. Goity, H. van Bakel, J. C. Lozano, M. Galli, M. G. Lewsey, E. Huang, T. Mukherjee, X. Chen, J. S. Reece-Hoyes, S. Govindarajan, G. Shaulsky, A. J. M. Walhout, F. Y. Bouget, G. Ratsch, L. F. Larrondo, J. R. Ecker and T. R. Hughes (2014). "Determination and inference of eukaryotic transcription factor sequence specificity." Cell **158**(6): 1431-1443.

Wu, C., X. Li, W. Yuan, G. Chen, A. Kilian, J. Li, C. Xu, X. Li, D. X. Zhou, S. Wang and Q. Zhang (2003). "Development of enhancer trap lines for functional analysis of the rice genome." Plant J **35**(3): 418-427.

Wu, C. H., Y. Yamaguchi, L. R. Benjamin, M. Horvat-Gordon, J. Washinsky, E. Enerly, J. Larsson, A. Lambertsson, H. Handa and D. Gilmour (2003). "NELF and DSIF cause promoter proximal pausing on the hsp70 promoter in Drosophila." Genes Dev **17**(11): 1402-1414.

Xu, H. and T. R. Hoover (2001). "Transcriptional regulation at a distance in bacteria." Curr Opin Microbiol **4**(2): 138-144.

Yan, W., D. Chen, J. Schumacher, D. Durantini, J. Engelhorn, M. Chen, C. C. Carles and K. Kaufmann (2019). "Dynamic control of enhancer activity drives stage-specific gene expression during flower morphogenesis." Nat Commun **10**(1): 1705.

Yang, W., R. A. Jefferson, E. Huttner, J. M. Moore, W. B. Gagliano and U. Grossniklaus (2005). "An egg apparatus-specific enhancer of Arabidopsis, identified by enhancer detection." Plant Physiol **139**(3): 1421-1432.

Yu, C. P., S. C. Chen, Y. M. Chang, W. Y. Liu, H. H. Lin, J. J. Lin, H. J. Chen, Y. J. Lu, Y. H. Wu, M. Y. Lu, C. H. Lu, A. C. Shih, M. S. Ku, S. H. Shiu, S. H. Wu and W. H. Li (2015). "Transcriptome dynamics of developing maize leaves and genomewide prediction of cis elements and their cognate transcription factors." Proc Natl Acad Sci U S A **112**(19): E2477-2486.

Yuan, G. C., Y. J. Liu, M. F. Dion, M. D. Slack, L. F. Wu, S. J. Altschuler and O. J. Rando (2005). "Genome-scale identification of nucleosome positions in S. cerevisiae." Science **309**(5734): 626-630.

Zabidi, M. A., C. D. Arnold, K. Schernhuber, M. Pagani, M. Rath, O. Frank and A. Stark (2015). "Enhancer-core-promoter specificity separates developmental and housekeeping gene regulation." Nature **518**(7540): 556-559.

Zentner, G. E. and P. C. Scacheri (2012). "The chromatin fingerprint of gene enhancer elements." J Biol Chem **287**(37): 30888-30896.

Zentner, G. E., P. J. Tesar and P. C. Scacheri (2011). "Epigenetic signatures distinguish multiple classes of enhancers with distinct cellular functions." Genome Res **21**(8): 1273-1283.

Zhang, P., J. H. Xia, J. Zhu, P. Gao, Y. J. Tian, M. Du, Y. C. Guo, S. Suleman, Q. Zhang, M. Kohli, L. S. Tillmans, S. N. Thibodeau, A. J. French, J. R. Cerhan, L. D. Wang, G. H. Wei and L. Wang (2018). "High-throughput screening of prostate cancer risk loci by single nucleotide polymorphisms sequencing." Nat Commun **9**(1): 2022.

Zhang, W., Y. Wu, J. C. Schnable, Z. Zeng, M. Freeling, G. E. Crawford and J. Jiang (2012). "High-resolution mapping of open chromatin in the rice genome." Genome Res **22**(1): 151-162.

Zhang, W., T. Zhang, Y. Wu and J. Jiang (2012). "Genome-wide identification of regulatory DNA elements and protein-binding footprints using signatures of open chromatin in Arabidopsis." Plant Cell **24**(7): 2719-2731.

Zhu, B., W. Zhang and T. Zhang (2015). "Genome-Wide Prediction and Validation of Intergenic Enhancers in Arabidopsis Using Open Chromatin Signatures." **27**(9): 2415-2426.