Mapping and dynamics of regulatory DNA during seed development

Alessandra M. Sullivan¹, Andrej A. Arsovski², Agnieszka Thompson¹, Richard Sandstrom¹, Robert E. Thurman¹, Shane Neph¹, Audra K. Johnson¹, Shawn T. Sullivan¹, Peter J. Sabo¹, Fidencio V. Neri III¹, Molly Weaver¹, Morgan Diegel¹, Jennifer L. Nemhauser², John A. Stamatoyannopoulos¹, Kerry L. Bubb¹,*, Christine Queitsch¹

¹ Dept. of Genome Sciences, University of Washington, Seattle, WA, 98195, USA
² Dept. of Biology, University of Washington, Seattle, WA, 98195, USA

*Correspondence: kbubb@uw.edu

Short title: Regulatory DNA dynamics in the maturing seed
Abstract

The genome is reprogrammed during development to produce diverse cell types, largely through altered expression and activity of key transcription factors. The accessibility and critical function of epidermal cells have made them a model for connecting transcriptional events to development in a range of model systems. In Arabidopsis and many other plants, fertilization triggers differentiation of specialized epidermal cells called the seed coat that have a unique morphology caused by large extracellular deposits of pectin. Here, we used DNase I-seq to generate regulatory landscapes of *Arabidopsis thaliana* seeds at two points in development. To enrich for seed coat signals, we used the INTACT method for capturing nuclei from GL2-expressing cells from whole siliques. Over half of the regulatory (i.e., accessible) bases identified in seeds at these two developmental time points had not been previously identified in seven-day-old seedling. We identified over 3000 regions that were developmentally dynamic (i.e., differentially accessible) at the two timepoints and found that genes neighboring these regions were enriched for gene ontology terms such as development, regulation and pigment. Genes neighboring differentially accessible regions were also significantly more likely to have differing expression at these two timepoints. The differentially accessible regions themselves were enriched for motifs belonging to transcription factors involved in development and environmental response, notably the TCP and MYB families at the earlier and later time points, respectively. These findings are consistent with previous studies, highlighting the dramatic changes that take place in the chromatin landscape during development, and indicating the importance of extending the approach described here to other cell types and developmental processes.

**Keywords:** DNase I-seq, regulatory DNA, *Arabidopsis thaliana*, seed development, open chromatin, accessible DNA
Introduction

Spatial and temporal regulation of gene expression is critical for development and specialization of tissues and cell types. Cis-regulatory DNA elements, and the trans-acting factors that bind them, are a primary mechanism for gene regulation. Active cis-regulatory elements such as promoters, enhancers, insulators, silencers, and locus control regions can be identified by their characteristic hypersensitivity to cleavage by DNase I (Banerji, Olson, and Schaffner 1983; Baniahmad et al. 1990; Chung, Bell, and Felsenfeld 1997; Talbot et al. 1989; Thurman et al. 2012; Wu et al. 1979; Wu, Wong, and Elgin 1979). Our previous analyses of regulatory DNA and its dynamics in Arabidopsis (A. thaliana) have focused on identifying regulatory modules in whole seedlings exposed to environmental cues (A. M. Sullivan et al. 2014). Our method, which relies on INTACT-labeled nuclei (Deal and Henikoff 2010), lends itself readily to investigating the regulatory landscape of individual cell types. Cell-type-enriched approaches to gene regulation and expression are fundamental for understanding development. Here, we use DNase I-seq to examine the regulatory landscape of seeds at two developmental timepoints, four and seven days post-anthesis (after the opening of the flower opening, or approximately one day after self-fertilization). Two main developmental changes occur between these two time points: maturation of the seed coat and rapid growth and differentiation of the embryo. Our method enriches for GL2-expressing cells. These are mainly the seed coat at both time points in addition to the embryo at the latter time point (Figure 1A).

The seed coat differentiates from the integuments of the ovule after fertilization has occurred. In many species, seed coat cells produce and store polysaccharide-rich mucilage (myxospermy). When wetted, this mucilage expands and extrudes from mucous-secreting cells, forming a gel-like layer around the seed (T. L. Western, Skinner, and Haughn 2000; Windsor et al. 2000). In A. thaliana, mucilage is composed primarily of pectin with lesser amounts of cellulose and xyloglucan (Haughn and Western 2012). Although the function of mucilage
depends on the species and the environmental context (García-Fayos, Bochet, and Cerdà 2010; Garwood 1985; Gutterman and Shem-Tov 1997; Yang, Dong, and Huang 2010; Yang et al. 2011), it is generally thought to protect the seedling and facilitate dispersal and germination.

In *A. thaliana*, seed coat cell differentiation and maturation is well characterized at the morphological level (T. L. Western, Skinner, and Haughn 2000; Windsor et al. 2000). In the mature ovule seed coat cells are characterized by a large vacuole. During the first four days after fertilization, the vacuole swells causing cell growth, and starch granules appear. By seven days after fertilization, the vacuole shrinks, the cytoplasm forms a column filled with vesicles and golgi stacks, and mucilage is secreted into the apoplast. By ten days post fertilization, mucilage production is complete, and a secondary cell wall is being deposited around the columnar cytoplasm forming a solid structure, the columella. Once differentiation is complete, dehydration shrinks the stored mucilage causing the primary cell wall to drape over the newly formed columella, creating the polygonal pattern visible on the dry seed exterior.

Seed coat cells are an exceptionally well-studied plant cell type. Previous studies have identified around 50 genes affecting seed coat cell differentiation and maturation when disrupted in *A. thaliana* (Francoz et al. 2015; North et al. 2014). These genes fall into roughly three categories: epidermal cell differentiation, mucilage synthesis and secretion, and secondary cell wall synthesis (*Supplemental Table 1*). Genes controlling specification of the ovule integument will also impact seed coat cell differentiation. Many of the genes required for seed coat differentiation and mucilage production are transcription factors (*Supplemental Table 1*) (Francoz et al. 2015).

While the identity of TFs, and in some cases their targets, are known, there is little information about individual regulatory elements and their activation during seed coat epidermis differentiation and maturation. Exceptions include the promoter of *DP1*, which specifically drives seed coat epidermal expression (Esfandiari et al. 2013), and the L1 box in the *CESA5*
promoter, which interacts with GL2 (a seed coat epidermis differentiation factor) in yeast (Tominaga-Wada et al. 2009).

To address this paucity of genome wide regulatory information, we employed the INTACT method to capture the nuclei of GL2-expressing cells from whole siliques, followed by DNase I-seq to identify regulatory elements, their dynamics, and their constituent TF motifs at two critical time points in seed development. We observe dramatic changes in the regulatory landscape, relate dynamic DNase I-hypersensitive sites (DHSs) to previously established expression profiles, identify genes that neighbor dynamic DHSs, and identify associated transcription factor motifs. We identify many candidate genes that may contribute to seed coat development in ways that might escape traditional genetic analysis.

By comparing our novel seed coat-enriched regulatory landscapes to previously generated landscapes we identified many novel DHSs. As in animals (Stergachis et al. 2013; Thomas et al. 2011; Daugherty et al. 2017), cell lineage and developmental stage are strong determinants of the chromatin landscape. A systematic analysis of A. thaliana cell types during development and in response to major environmental cues will be especially fruitful in discovering additional, developmentally-important regulatory elements.

Results

The regulatory DNA landscape of maturing seed coat epidermal cells

To capture the regulatory landscape of seed coat epidermal cells, we employed nuclear capture (INTACT) (Deal and Henikoff 2010) followed by DNase I-seq (A. M. Sullivan et al. 2014). We used an existing transgenic plant line (Deal and Henikoff 2010) in which the GL2 promoter controls the targeting of biotin to the nuclear envelope (Supplemental Figure 1). GL2
is consistently expressed at high levels in the seed coat epidermis, but is also expressed to varying degrees elsewhere in the seed, most noticeably in the embryo (Windsor et al. 2000; Belmonte et al. 2013). We sampled whole siliques, which encase 40-60 seeds, at 4 and 7 days post-anthesis (DPA), to capture the regulatory landscape before and after mucilage production begins in the seed coat.

We created five DNase I-seq libraries, including biological replicates for each time point, and identified a union set of 43,120 DHSs. Of these DHSs, 3,109 were determined to be developmentally dynamic between 4DPA and 7DPA by DEseq2 (Love, Huber, and Anders 2014) with an adjusted p-value of < 0.001 (Figure 1A, Supplemental Tables 2-5, Methods).

Twenty dynamic DHSs neighbor one of 48 known seed coat development genes (Supplemental Table 1). For example, we found 7DPA-activated DHSs near MYB61, which is required for mucilage production (Penfield et al. 2001), and PER36, which is required for proper mucilage release (Kunieda et al. 2008) (Figure 1B). We also identified many dynamic DHSs near genes that were not previously associated with seed coat development. For example, the meristem identity transition TF, LMI2 (Pastore et al. 2011), was near a DHS that is deactivated during seed coat cell maturation (Figure 1B). Similar to previous observations (A. M. Sullivan et al. 2014), the majority of observed DHSs were static during development, such as those flanking CES5, which encodes a cellulose synthase that produces seed mucilage cellulose (S. Sullivan et al. 2011) (Figure 1B). The landscape of seed coat cells differed significantly from the regulatory landscape of another epidermal cell type, root nonhair cells, as well as whole roots (Figure 1B, Figure 5). We conclude our method detects developmentally regulated seed coat cell DHSs, many of which appear in the vicinity of known seed coat development genes.

We also observed that the developmentally dynamic DHSs appeared to be clustered. Each dDHS has two neighbors, so in theory, there could be a maximum of 6,218 genes neighboring
the 3,109 dDHSs. Only 4,791 of those genes are unique, however, since many neighbor more than one dDHS (Figure 1C).

Next, we asked if the genomic distribution of dynamic DHSs was different than that of all DHSs by tabulating the number of DHSs occurring in various genomic contexts (e.g. intragenic) (Supplemental Table 6). Similar to whole seedling DHSs (A. M. Sullivan et al. 2014), DHSs in GL2-expressing cells (both dynamic and static), are enriched in intergenic regions and regions 400 bps upstream of transcription start sites (labeled TSS region), and are depleted in intragenic regions and transposable elements (TEs). Developmentally dynamic DHSs were primarily enriched in intergenic regions (Figure 1D). This distribution is consistent with previous observations in Drosophila, where developmental enhancers are primarily located in intergenic regions and in introns while housekeeping gene enhancers are primarily located near transcription start sites (Zabidi et al. 2015).

**Genes neighboring dynamic DHSs are enriched for differentially expressed genes**

Of the 28,775 annotated genes in TAIR10, 4,791 (16.6%) neighbor one or more of the 3,109 developmentally dynamic DHSs, with some genes flanked by as many as ten developmentally dynamic DHS (Figure 1C). We hypothesized, that developmentally dynamic DHSs would be associated with genes that are differentially expressed during seed coat development. To address this hypothesis, we took advantage of two published seed coat epidermis expression studies (Belmonte et al. 2013; G. Dean et al. 2011), considering a gene to be differentially expressed if it exhibited a 2-fold expression change between developmental time points. We also included a set of 48 genes identified through a literature search to be involved in seed coat development (Supplemental Table 1).
In the first study, Dean et al. 2011 quantified gene expression in manually dissected seed coats at 3DPA and 7DPA in the Col-2 accession, identifying 3,430 genes that exhibited at least a 2-fold expression change between these timepoints. Seven of these genes appeared in both the up in 7DPA and up in 3DPA sets on different types of arrays and were excluded from our analysis, leaving 3,423 genes. In the second study, Belmonte et al. 2013 quantified gene expression in many parts of the seed at many time points in the Ws-0 ecotype using laser capture micro dissection. For our analysis, we used the seed coat (SC) and embryo proper (EP) expression values from globular (g), heart (h) and linear cotyledon (lc) stage seeds, which approximate the 3-4DPA, 4-5DPA and 7DPA stages, respectively (Le et al. 2010). Both studies used microarrays to evaluate gene expression.

We found a 2.5-fold enrichment of the 48 genes known to be involved in seed development among the set of unique genes neighboring one or more dynamic DHS than expected by chance, and 1.63 and 1.60 -fold enrichments of that genes with differing expression in seed coat between globular and linear cotyledon stage seeds in the Belmonte and Dean sets, respectively (Figure 2A). Furthermore, increased accessibility was significantly associated with increased expression level at both the 4DPA and 7DPA stages (Figure 2C). Conversely, decreased accessibility was also associated with lower expression level, however this association was not statistically significant (Figure 2C). Although 4DPA seeds are mainly in the globular phase of development, some 4DPA seeds have progressed to the heart stage. It is known that the INTACT transgene promoter (GL2) is activated in the embryo of both heart and linear cotyledon stage seeds. We therefore also examined the enrichment of genes differentially expressed between heart and linear cotyledon stage seeds in seed coat and in embryo (Supplemental Figure 2A&B). As with the globular vs linear cotyledon comparison, there was a 1.58-fold enrichment of genes differentially expressed in the seed coat in heart vs linear cotyledon among the set of genes neighboring one or more dynamic DHS, whereas genes differentially expressed in the embryo between these stages were only slightly, albeit significantly, enriched in the
dDHS-neighboring set (1.17-fold). An example of a gene with expression differences in embryo at different stages of development but not in seed coat is cytochrome P450 (AT1G13710). This gene has higher expression in heart staged embryos than linear cotyledon staged embryos, and it is flanked by two differential DHSs that are both more accessible in 4DPA than 7DPA.

We next explored whether genes neighboring multiple dynamic DHSs were more enriched in our known, Belmonte, and Dean gene sets. Indeed, there was a monotonic increase in fold-enrichment for each of these three data sets when examining genes neighboring one or more, two or more, or three or more dynamic DHS (Supplemental Figure 2C).

**Genes nearest to dynamic DHSs are involved with seed coat biology**

To test whether the genes that resided nearest to dynamic DHSs (Figure 2A) were involved in known seed coat epidermis biology, we analyzed their GO terms using GOstats (Figure 3; Supplemental Tables 7&8). Genes nearest to deactivated DHSs (more accessible in 4DPA seeds than 7DPA seeds) were enriched for development, regulation, response, and pigment genes. Genes nearest to activated DHSs (more accessible in 7DPA seeds than 4DPA seeds) were enriched in genes related to transport, cell wall, biosynthetic process, and localization. The twenty genes overlapping the set of 48 genes known to be involved in seed coat development are also listed in Figure 3C.

We identified likely seed coat differentiation and maturation candidate genes based on their expression and/or chromatin accessibility profiles, and tested confirmed homozygous T-DNA mutants for mucilage defects as previously described (Western et al., 2000) (Supplemental Table 13). However, none of these mutants showed obvious defects in mucilage production.
Motif families in activated and deactivated DHSs are distinct

To explore the drivers of deactivated and activated DHSs in seed coat development, we looked for motif enrichment in developmentally dynamic DHSs as compared to union DHSs using AME in the meme-suite (McLeay and Bailey 2010). We find that different motifs are enriched at different timepoints. For example, bHLH and TCP motifs are enriched in DHSs with greater accessibility at 4DPA than 7DPA (Figure 4A), while many more motif families are enriched in DHSs with greater accessibility at 7DPA than 4DPA, including ARID, bZIP, MADS, MYB, MYB-related, and NAC, with the MYB family having the greatest enrichment by far (Figure 4B).

Comparative analysis of diverse plant regulatory landscapes

Previous studies in humans comparing regulatory landscapes of many cell types revealed cell lineage is encoded in the accessible regulatory landscape (Stergachis et al. 2013). Similarly, a dendrogram generated using accessibility profiles generated from thirteen plant samples primarily reflects ontogeny, with only subtle changes brought about by treatments (Figure 5A). For example, seven-day old seedlings germinated and grown in the dark inhabit a distinct clade from seedlings grown in the typical light cycle. On a finer scale, seedlings exposed to heat near time of harvest cluster with other seedlings grown in the typical light cycle, and dark-grown seedlings exposed to the light-mimicking hormone brassinazole (BRZ) cluster with seedlings exposed to light near time of harvest. This is also evident in a Principal Component Analysis biplot, showing the sample vectors projected on the PC1-PC2 plane (Figure 5B). It should be noted that the silique samples (4DPA and 7DPA) are derived from adult plants, more than 20 days old. Nevertheless, this result is consistent with a meta study indicating that expression profiles are more different between tissue types than between tissue-controlled treatments (Aceituno et al. 2008).
We then examined which of the thirteen samples provided the most novel hypersensitive bases. Merging DHSs from each of these thirteen samples generates 46,891 union DHSs, covering 10,374,430 bases of the genome (see Methods for details). If we exclude each of the thirteen samples individually, hypersensitive bases unique to that sample are lost. The silique samples (4DPA and 7DPA) contribute the most unique hypersensitive bases, followed by the root of seven-day old seedling (Figure 6). Furthermore, of the hypersensitive bases identified in the silique samples, over half (2,858,990 bps / 5,573,620 bps) were not present in 7-day-old seedlings, and over 25% (1,418,070 bps / 5,573,620 bps) were not present in any of the other eleven samples examined.

Finally, as these samples were initially prepared with pairwise comparisons in mind, we determined which pairs provided the most developmentally dynamic sites (Figure 7). We found that the 4DPA vs 7DPA comparison revealed many more developmentally dynamic regions than any other pair, with heat-shock vs control a far second, and root hair vs root non-hair having the lowest number of developmentally dynamic sites. For this analysis, we defined developmentally dynamic DHSs as those that lie above or below a cone and, because we plotted all union DHSs regardless of whether they were identified as DHSs in the samples compared, we excluded DHSs that had very low accessibility in both samples (see Methods).

**Discussion**

Here, we mapped regulatory elements and their developmental dynamics in GL2-expressing cells from whole siliques using DNase I-seq. We targeted the developmental stages in which the seed coat transitions from a state of growth to a state of mucous production and secretion. During this developmental window, more than 3,000 DHSs change in accessibility.
DHSs are a hallmark of regulatory DNA and thus dynamic DHSs often neighbor genes with changing expression. We found that genes that are differentially expressed during seed coat cell development often neighbor dynamic DHSs. It is known that the association between increasing accessibility and increasing expression of neighboring genes is imperfect for several reasons. First, regulatory DNA can be poised (i.e. accessible) for transcription activation (Elgin 1988) and DHSs can remain accessible after transcription is shut off (Groudine and Weintraub 1982). Second, the binding of activators (Morgan et al. 1987) or repressors (Banahmad et al. 1990) has the potential to remodel chromatin.

Despite these limitations, dynamic DHSs are potentially useful in identifying new candidate genes that control seed coat development, and suggesting TFs that may be driving or responding to DHS dynamics. Genes near 4DPA-activated DHSs were often associated with development, signaling, pigment, and regulation, consistent with changes that occur during the seed maturation. Genes near 7DPA-activated DHSs were often associated with the transport, localization, biosynthetic processes, and cell wall modification, consistent with these cells switching to mucous production and secretion into the apoplast, and ramping up to build the columella, a secondary cell wall structure. Motif enrichments within 7DPA-activated and 7DPA-deactivated DHSs revealed distinct TF families and individual TFs that may be regulating seed coat cell maturation. Previous functional studies validate our motif findings.

Among motifs most enriched in 7DPA-deactivated DHSs were those of the TCP family. TCPs are known to be involved in many aspects of development, particularly in land plants in which the class has greatly diversified (Martín-Trillo and Cubas 2010). Overexpression of TCP3, for example, leads to ovule integument growth defects and ovule abortion (Wei et al. 2015). Indeed, it was once thought that TCPs originated in land plants, however more thorough analysis revealed the presence of TCPs in the freshwater-dwelling genus Chara, a type of multicellular algae with stem- and leaf-like structures, indicating that common ancestor of TCPs predates the
origin of land plants (Navaud et al. 2007). It is therefore fitting that one of the two original sequences in the family -- and arguably the most famous member -- is TB1, the gene responsible for major morphological changes during the evolution of teosinte to maize (Cubas et al. 1999; Clark et al. 2006)

Among motifs most enriched in 7DPA-activated DHSs were those of the MYB family. This class of TFs is represented throughout the eukaryotic taxon, but is particularly involved in development and stress-response in plants (Ambawat et al. 2013). All of the MYB TFs with enriched motifs in 7DPA DHSs belonged to the same subfamily, the R2R3 MYBs, which are involved in secondary metabolism and cell fate establishment (Stracke, Werber, and Weisshaar 2001). For example, MYB2 is known to induce seed-trichome production (Wang et al. 2004), and MYB5 and MYB61 are known to be important for mucilage production (Penfield et al. 2001; Li et al. 2009; Gonzalez et al. 2009). Zinc finger, MADS-box, and AT-hook TFs were also enriched in 7DPA-activated DHSs; these TF families have not been implicated previously in seed coat cell maturation. However, MADS-box TFs are required for proper ovule development (Honma and Goto 2001; Pinyopich et al. 2003).

While none of the mutants tested showed obvious defects in mucilage production, phenotypes are notoriously difficult to pinpoint. For example, a recent study (Khosla et al., 2014) showed that hdg11 mutants have normal seed coat mucilage. However, when HDG11 is expressed under the control of the GL2 promoter (which causes strong expression in the seed coat epidermis) in the hdg11 background, seed coat mucilage production is inhibited in a dominant negative fashion. This example illustrates the importance of tight control over HDG11 expression for normal seed coat development, and more importantly, suggests that the best way to study seed coat mucilage regulators may be through the targeted disruption of individual regulatory elements rather than through the analysis of mutant lines. Indeed, another study using gene expression data to pinpoint mucilage regulators similarly yielded negative results in single
mutant analysis (Dean et al., 2011). Future experiments are needed to explore the function of individual regulatory elements through targeted disruption, the potential combinatorial action of candidate genes, and subtler mutant phenotypes related to seed coat function.

This foray into cell-type-specific regulatory landscapes in plants, demonstrates the significant knowledge gains by analyzing specific cell types and their developmental dynamics rather than using whole seedlings or easily dissected tissues. The regulatory elements identified in this study can now be integrated with the existing co-expression- and genetics-based gene regulatory network data to gain a more complete understanding of the regulation of seed coat maturation (Francoz et al. 2015).

Acknowledgements

This work was supported by grants from the National Science Foundation (MCB1243627) (JS, CQ, JN), and Graduate Research Fellowship (DGE-0718124) (AMS). We thank Roger Deal and Steven Henikoff for sharing INTACT lines and experimental expertise, members of the Stamatoyannopoulos and Queitsch labs for useful discussions. We thank Tamara Western for sharing her mucilage staining protocol and for giving us seeds for known mucilage mutants (ttg2-1, bxl1, and mum4-1), and controls (Col-0 and Col-2). We thank Chris Gee for technical assistance with DNase I preps, and G. Alex Mason for drawing and colorizing the seeds at different stages of development. The authors have no conflicts of interest.

Author Contributions

J.A.S. and J.L.N facilitated experiments and assisted in writing the manuscript. All authors read, commented on, and approved the manuscript.

Methods

Sample preparation

Siliques of appropriate ages from the INTACT line $GL2_{pro}:NTF/ACT2_{pro}: BirA$ (Deal and Henikoff 2010) were collected by first marking young flowers using a fine paint brush and water based paint as previously described (T. L. Western, Skinner, and Haughn 2000). In brief, recently opened flowers are chosen at the stage the anthers are almost at the same level as the pistil and fertilization is able to occur, usually 2 per plant per day at this stage. The flower is marked with paint and siliques collected 4 or 7 days later. Samples were prepared using INTACT nuclei isolation (Deal and Henikoff 2010) followed by DNase I-seq (A. M. Sullivan et al. 2014). A detailed protocol for tissue preparation and nuclei isolation using INTACT lines is provided at plantregulome.org. A detailed protocol for post-digestion sample processing has been published previously (John et al. 2013). Data sets may be found in GEO accessions GSE53322 and GSE53324 and at plantregulome.org.

Microscopy

Testing activity of the INTACT construct in seed coat cells

Whole seeds were observed on a Leica TCS SP5 II laser scanning confocal microscope. Whole seed images (Supplemental Figure 1A) are z-stack composites of 35 individual images using an HC Plan Apo CS 20X objective. Image of seed coat cell layer (Supplemental Figure 1B) is a single image using the 63X water immersion objective.
Candidate screening

Seeds from mutants (candidates) as well as controls (Col-2, ttg2-1, bxl1, and mum4-1) were imbibed with water for 1 hour with shaking at 200 rpm according to (T. L. Western, Skinner, and Haughn 2000). Seeds were then stained with 0.01% Ruthenium red (SIGMA) for 1 hour. Seeds were washed and mounted on a slide and viewed with a light microscope.

Data processing for seed coat analysis

Five DNase I-seq libraries, including biological replicates for each time point, were sequenced. Because number of peaks called is a function of read depth, 24 million reads mapping to chromosomes 1-5, excluding centromeres (chr1:13,698,788-15,897,560; chr2: 2,450,003-5,500,000; chr3:11,298,763-14,289,014; chr4:1,800,002-5,150,000; chr5:10,999,996-13,332,770), were sampled from the biological replicate with the highest read coverage for each developmental time point (4DPA-DS20201 and 7DPA-DS21306). These 24M-read bam files were used to call DHSs (peaks) using the HOTSPOT program (John et al. 2011a). DHSs from these two samples were merged to create a union set of 43,120 DHSs. DESeq2 (Love, Huber, and Anders 2014) was used on this set of union DHSs to identify a subset of 3,440 developmentally dynamic DHSs (adjusted p-value < 0.01), using all reads mapping to chromosomes 1-5, excluding centromeres, from all five samples (4DPA-DS20201, 4DPA-DS20131, 4DPA-DS20132, 7DPA-DS21306, 7DPA-DS20134). We then removed DHSs with mean cut count of 50 or less -- roughly the bottom ten percentile -- leaving 3,109 dynamic DHSs. Data sets may be found in GEO accessions GSE53322 and GSE53324 and at plantregulome.org.

Genomic distribution of DHSs

DHS midpoints were used to determine overlaps with genomic elements. Genomic elements (5’UTR, coding regions, 3’UTR, intergenic, TE) were extracted from the TAIR10 gff file on
arabidopsis.org. Centromeric regions were excluded from the analysis. To simplify the analysis, only the primary transcript of each gene (AT*.1) was considered. When a single DHS midpoint coincided with two different elements, both element overlaps were tallied, thus overlapping DHS counts sum to greater than the initial number of DHSs. We tallied the total number of base pairs within each element type in the genome, double-counting base pairs that are assigned to overlapping elements. Tallies may be found in Supplemental Table 6.

**Integration with expression data sets**

Genes from Dean et al. 2011 and Belmonte et al. 2013 were considered to be differentially expressed if there was a 2-fold change in expression between time points. Dean et al. 2011 identify the genes that change 2-fold between 3DPA and 7DPA; these genes were used for integration with dynamic DHS data. The genes that change expression by 2 or more fold in Belmonte et al. 2013 were extracted from the published normalized expression data (Dataset S2).

**Term enrichment**

Term enrichments were performed using the org.At.tair.db (Carlson 2016) and GOstats (Falcon and Gentleman 2007). Only the enrichments with a p-value less than 0.001 are shown in Figure 3.

**Motif enrichment**

Enrichment of motifs (O’Malley et al. 2016) in sequence underlying dDHSs as compared to union DHSs was evaluated using AME (McLeay and Bailey 2010). All members of motif families in which at least one member is enriched with significance of $p<10^{-20}$ are displayed in Figure 4. All motifs with corrected p-value of less than 0.01 are listed in Supplemental Tables.
Motifs derived using amplified DNA (colamp_a) are gray and motifs derived using native genomic DNA (col_a) are black.

**Comparative analysis of DHS landscapes**

Each of 13 samples was subsampled to roughly 14 million reads mapping to chromosomes 1-5, excluding centromeres (chr1:13,698,788-15,897,560; chr2: 2,450,003-5,500,000; chr3:11,298,763-14,289,014; chr4:1,800,002-5,150,000, chr5:10,999,996-13,332,770) (**Supplemental Table 11**). DHSs were called on these 13 bam files using the HOTSPOT program (John et al. 2011b), and a union set of DHSs was generated by merging DHSs from each of these 13 samples with BEDOPS (Neph et al. 2012), (bedops –m, adding each sample in succession) (**Supplemental Table 12**). There were 62,738 DHSs in this union set. Per-base DNase I cleavages (cut counts) within each union DHS were tallied for each sample. Cleavage tallies were normalized for sample quality by dividing by the proportion of DNase I cleavages within 1% FDR threshold hotspots.

**Accessibility profiles used to cluster samples**

Dendrogram and bootstrap values were generated 100 trees from random subsamples of 10,000 DHSs using the ape package (Paradis, Claude, and Strimmer 2004). Principal Component Analysis was performed on the 62,729 by 13 matrix. For the PCA, we excluded nine DHSs within the first 50kb of chromosome 2, part of a NOR (nucleolar organizer region) (Copenhaver and Pikaard 1996; Lin et al. 1999), a region with unusually high cut count, similar to the centromeres.

**Sample-specific hypersensitive bases**

To identify sample-specific hypersensitive bases, we merged large DHSs (>50 cleavages per DHS) from the 13 samples to generate a set of 46,891 union DHS covering 10,374,430 bps. We
then generated 13 new merged sets of DHSs using only 12 samples, excluding one of the samples in each set, and then determined the number of hypersensitive bases not captured. We define the number of hypersensitive bps unique to the sample as number of bps in the 13-sample union DHS set minus the number of bp in the 12-sample union DHS set divided by the number of bps in the 13-sample union DHS set (Figure 6).

**Pairs of samples resulting in differential DHSs**

To compare the number of developmentally dynamic DHSs identified with different pairs of samples, we used the complete set of merged DHSs (62,738 unionpeaks). For each of six pairwise comparisons, we made a scatterplot of the cut counts of these 62,738 unionpeaks. We then defined developmentally dynamic DHSs as those that both lie outside a cone defined by the lines $y=(1-0.21)x + 0.9$ and $y=(1+0.21)x - 0.9$ and have greater than 50 cleavages per unionpeak in at least one sample. Expression differences between these pairs have been previously published (A. M. Sullivan et al. 2014).

**Candidate gene selection**

Candidates were identified because they were nearest to 7DPA-activated or –deactivated DHSs (in some cases multiple dynamic DHSs), and in all but one case, because they had been previously identified as differentially expressed in seed coat cells (G. Dean et al. 2011; Belmonte et al. 2013). BZO2H2 was chosen because it is a TF that is upregulated during seed coat cell differentiation. RD26 was chosen because it has many DHSs in its promoter proximal region and is in close proximity to a dynamic DHS. If possible, exonic SALK lines were obtained; only confirmed homozygous SALK lines were screened (Alonso et al. 2003). Candidates are located in Supplemental Table 13.
Data Access

**Figure 1. The chromatin landscape of maturing seed coat epidermal cells.**

A, Distribution of log2(DNase I cut count in 7DPA / DNase I cut count in 4DPA) for all union DHSs (gray) and differential DHSs, with DHSs more accessible at 4DPA appearing on the left in blue and DHSs more accessible at 7DPA appearing on the right in pink. Diagrams of 4DPA (left) and 7DPA (right) are shown, with purple opacity indicating expression level of GL2 from Belmonte et al. 2013. B, Examples showing a 7DPA-deactivated DHS, two examples of 7DPA-activated DHSs, and one example of a static DHS. A 5kb region is shown in each window; all data tracks are read-depth normalized. C, Distribution of the number of genes neighboring 1-10 dynamic DHS. D, The numbers of union DHSs (uDHSs) and dynamic DHSs (dDHSs) within each genomic context: TSS, intergenic, transposon, and intragenic.
A

<table>
<thead>
<tr>
<th>gene set</th>
<th>total</th>
<th>neighboring dDHS</th>
<th>fold</th>
</tr>
</thead>
<tbody>
<tr>
<td>Known</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>48</td>
<td>8</td>
<td>20</td>
</tr>
</tbody>
</table>

\[\Delta\text{Expr, 4DPA vs 7DPA}\]
Belmonte et al. 2013

|          | 4115  | 685             | 1120 | 1.63** |

Dean et al. 2011

|          | 3423  | 570             | 913  | 1.60** |

B

- genes changing expression, 4DPA vs 7DPA (Dean et al 2011) (n=3,423)
- genes changing expression, 4DPA vs 7DPA (Belmonte et al 2013) (n=4,115)

C

<table>
<thead>
<tr>
<th>genes with greater expression at 4DPA</th>
<th>neighboring dDHS more accessible in 4DPA than 7DPA</th>
<th>fold</th>
<th>neighboring dDHS more accessible in 7DPA than 4DPA</th>
<th>fold</th>
</tr>
</thead>
<tbody>
<tr>
<td>Belmonte et al. 2013</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1984</td>
<td>122 &lt;&lt; 289</td>
<td>2.37**</td>
<td>225 &gt;&gt; 182</td>
<td>0.61*</td>
</tr>
<tr>
<td>Dean et al. 2011</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1391</td>
<td>85 &lt;&lt; 166</td>
<td>1.94**</td>
<td>158 &gt; 135</td>
<td>0.85</td>
</tr>
</tbody>
</table>

4DPA

7DPA
Figure 2. Genes neighboring developmentally dynamic DHSs are often differentially expressed. **A**, Overlap between the set of genes neighboring dDHSs and 48 known seed coat development genes and genes differentially expressed in seed coat (G. Dean et al. 2011; Belmonte et al. 2013). **B**, Overlap of all four sets of genes. **C**, Genes that are more highly expressed tend to be near more accessible DHSs and vice versa. One asterisk (*) indicates p-value < 0.01. Two asterisks (**) indicate p-value < $10^{-20}$. 
Figure 3. Term enrichment for genes nearest to dynamic DHSs. A, Term enrichment for genes near DHSs that are less (left) and more (right) accessible at the 7DPA time point. B, The twenty genes of the 48 known seed coat development genes that neighbor one or more dynamic DHS.
**Figure 4. Motif enrichments within dynamic DHSs.** **A,** Motifs enriched in DHSs that are more accessible at the 4DPA timepoint. **B,** Motifs enriched in DHSs that are more accessible at the 7DPA timepoint. Dotted vertical line indicates adjusted p-value of $10^{-20}$ of $10^{-40}$. All family members are displayed if at least one member is enriched with adjusted p-value of $10^{-20}$ or less (greater than $-\log_{10}(10^{-20})$ or 20). Motifs derived using amplified (i.e., non-methylated) DNA have gray bars indicating enrichment p-value. Motifs derived from genomic (i.e., methylated) DNA have black bars indicating enrichment p-value.
Figure 5. Comparative analysis of diverse DHS landscapes. A, Dendrogram of thirteen samples using accessibility data. B, Biplot of Principal Component Analysis of 62,729-DHS by 13-sample matrix. Numbers in gray represent union DHSs (rows in the matrix). The accessibility of all 13 samples are displayed for two uDHSs that were highly informative for distinguishing samples (i.e. markedly differentially accessible).
Figure 6. Silique-derived samples contribute the most novel hypersensitive bases of the thirteen samples examined. Shown are the fraction of bps in union DHSs lost if a given sample is excluded from the union set.
Figure 7. Comparison of silique samples (4DPA and 7DPA) results in the highest number of developmentally dynamic DHSs identified among all pairs examined. Scatterplots of log10(cut counts per union DHS) for six pairwise comparisons. Dotted lines creating a cone capturing the majority of the dots are drawn in the same location on each graph. Gray boxes represents regions in which both samples have less than 50 \([\log_{10}(50)=1.69897]\) cleavage sites in that DHS. Numbers indicated above and below indicate the number of dots (DHSs) that lie above and below the line. Screenshot insets in each graph illustrating an example differential DHSs above and below the cone are the following, respectively: {4DPA vs 7DPA: chr2:19,564,381-19,564,531, chr4:11,981,161-11,981,351; root hair vs root nonhair: chr1:30,035,761-30,036,071, chr4:280,861-281,131; control vs auxin-treated: chr1:10,320,801-10,321,131, chr1:5,204,361-5,204,551; dark-grown seedling vs dark-grown seedling on BRZ: chr5:22,570,821-22,571,231, chr5:21,869,241-21,869,591; control vs heatshocked seedling: chr4:7,338,681-7,342,041, chr2:18,374,201-18,374,371; dark-grown seedling vs dark-grown seedling exposed to 24hr light cycle: chr3:6,023,601-6,023,871, chr5:5,968,041-5,968,291}
References


Haughn, George W., and Tamara L. Western. 2012. “Arabidopsis Seed Coat Mucilage Is a...
Specialized Cell Wall That Can Be Used as a Model for Genetic Analysis of Plant Cell Wall Structure and Function.” *Frontiers in Plant Science* 3 (April):64.


Kunieda, Tadashi, Tomoo Shimada, Maki Kondo, Mikio Nishimura, Kazuhiko Nishitani, and


Navaud, Olivier, Patrick Dabos, Elodie Carnus, Dominique Tremousaygue, and Christine Hervé. 2007. “TCP Transcription Factors Predate the Emergence of Land Plants.” *Journal of*


Sagasser, Martin, Gui-Hua Lu, Klaus Hahlbrock, and Bernd Weisshaar. 2002. “A. Thaliana
TRANSPARENT TESTA 1 Is Involved in Seed Coat Development and Defines the WIP Subfamily of Plant Zinc Finger Proteins.” *Genes & Development* 16 (1):138–49.


Yu, Li, Dachuan Shi, Junling Li, Yingzhen Kong, Yanchong Yu, Guohua Chai, Ruibo Hu, Juan...
Supplemental Figure 1. Confocal microscopy of INTACT-tagged nuclei in seed coat epidermis. 

A. Confocal of whole seed at 4DPA from the INTACT line $GL2_{pro}$:$NTF/ACT2_{pro}$:$BirA$ (Deal and Henikoff, 2010). GFP-fluorescing nuclei are evident across the seed coat epidermis. Scale is 100um.

B. Confocal of 4DPA mucous secreting cells (MSCs) from the INTACT line $GL2_{pro}$:$NTF/ACT2_{pro}$:$BirA$ (Deal and Henikoff, 2010). GFP-fluorescing nuclei are readily observable in the outer most layer of the seed coat. Scale is 100um.
Supplemental Figure 2. Genes neighboring developmentally dynamic DHSs are often differentially expressed in seed coat and embryo. A, Overlap between the set of genes neighboring dDHSs and genes differentially expressed in seed coat at globular vs linear
cotyledon stage and heart vs linear cotyledon stage, and genes differentially expressed in embryo at heart vs linear cotyledon stage (Belmonte et al. 2013). One asterisk (*) indicates p-value < 0.01. Two asterisks (**) indicate p-value < $10^{-20}$. B, Overlap of all four sets of genes. C, Genes neighboring multiple dynamic DHSs tend to be more enriched for seed coat development genes. This is seen in the set of 48 known seed coat development genes (Supplemental Table 1) as well as in genes with differential expression (G. Dean et al. 2011; Belmonte et al. 2013).