Mapping and dynamics of regulatory DNA during seed development

Alessandra M. Sullivan[1], Andrej A. Arsovski[2], Agnieszka Thompson[1], Richard Sandstrom[1], Robert E. Thurman[1], Shane Neph[1], Audra K. Johnson[1], Shawn T. Sullivan[1], Peter J. Sabo[1], Fidencio V. Neri III[1], Molly Weaver[1], Morgan Diegel[1], Jennifer L. Nemhauser[2], John A. Stamatoyannopoulos[1], Kerry L. Bubb[1,#], Christine Queitsch[1]

[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 *CESA5*, 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 (Baniahmad 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

A.A.A. designed the experiments, and together with A.T. executed DNase I-seq experiments. A.M.S. and K.L.B performed data analysis and made the figures. K.L.B., R.S, R.E.T, S.N., A.K.J. and S.T.S assisted with the bioinformatic analysis and data processing. P.J.S, F.V.N, M.W and M.D. sequenced DNase I-treated samples. A.M.S., K.L.B. and C.Q. wrote the manuscript.

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 silique 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**

9&10. 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

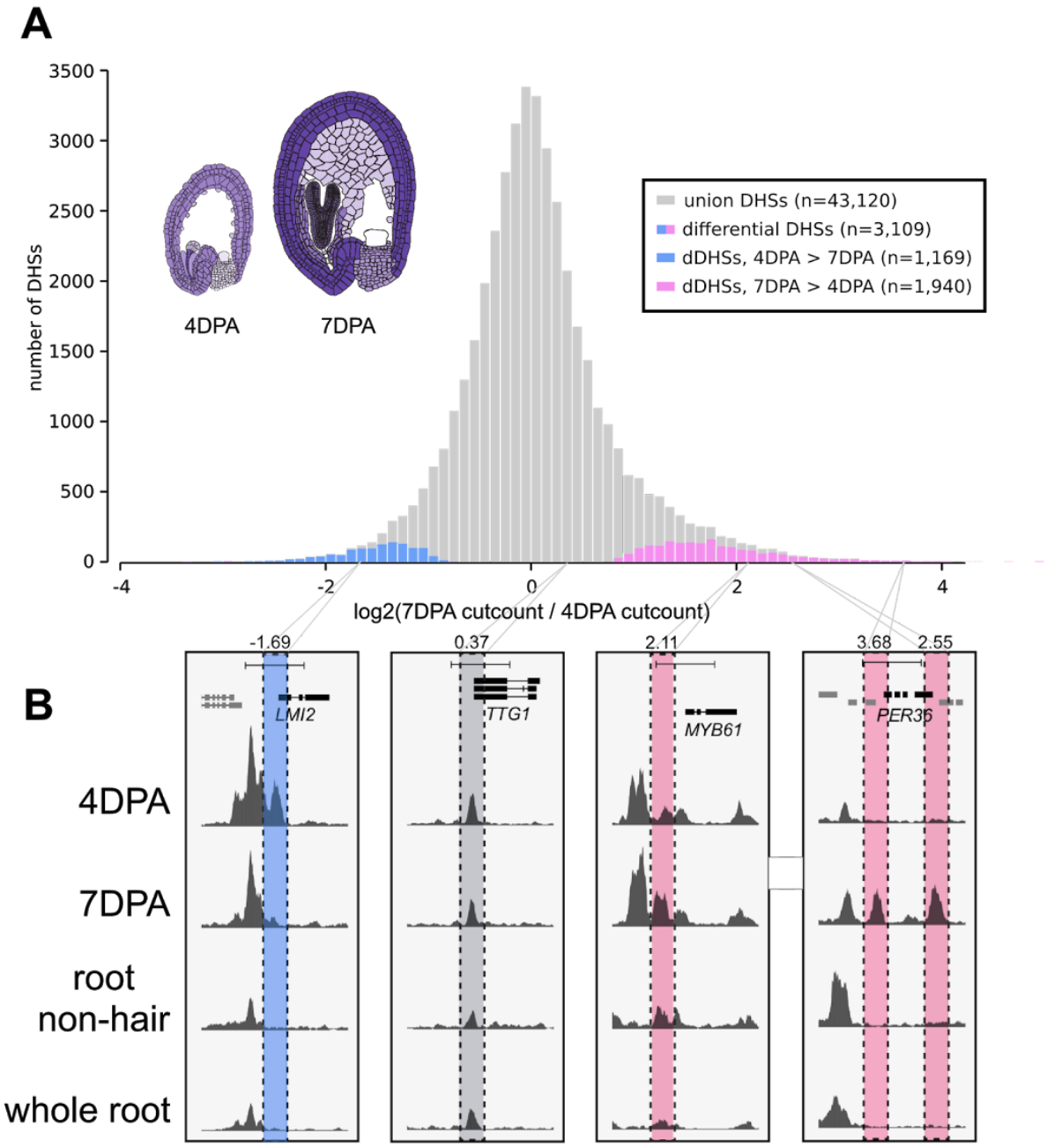
To compare the number of developmentally dynamic DHSs identified with different pairs of samples, we used the complete set of merged DHSs (62,738 unionpeaks). For each of six pairwise comparisons, we made a scatterplot of the cut counts of these 62,738 unionpeaks. We then defined developmentally dynamic 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

DNase I-seq data are available at http://www.plantregulome.org/data/releases and at GEO (https://www.ncbi.nlm.nih.gov/sra/; 4DPA-DS20201: SRR5873456; 4DPA-DS20131: SRR5873454; 4DPA-DS20132: SRR5873455; 7DPA-DS21306: SRR5873453 and 7DPA-DS20134: SRR5873452).



C

number of genes	number of dDHSs neighboring the gene
3,438	1
897	2
275	3
128	4
31	5
11	6
5	7
3	8
0	9
3	10

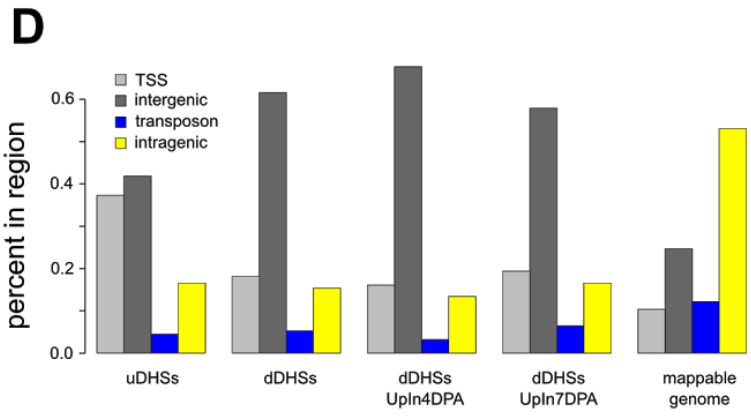
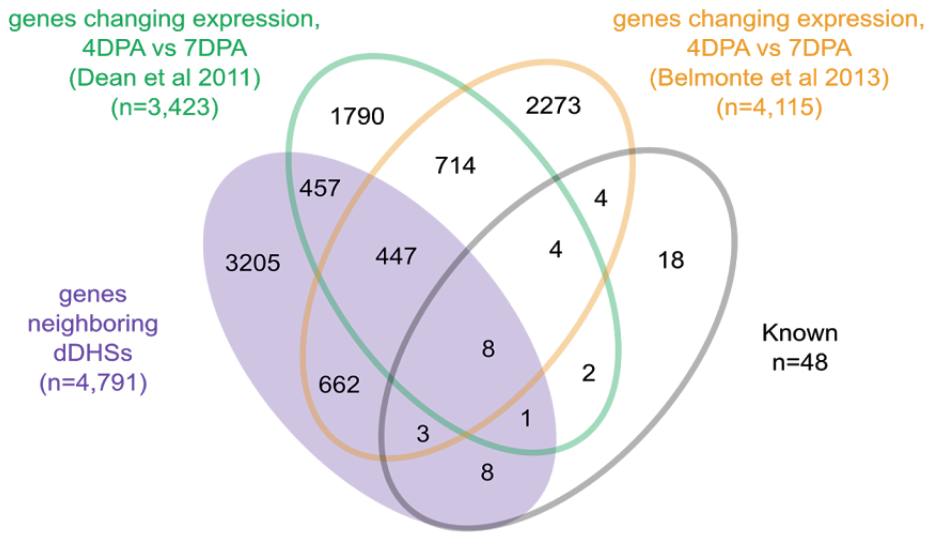


Figure 1. The chromatin landscape of maturing seed coat epidermal cells. A, Distribution of log2(DNase I cut count in 7DPA / DNase I cut count in 4DPA) for all union DHSs (gray) and differential DHSs, with DHSs more accessible at 4DPA appearing on the left in blue and DHSs more accessible at 7DPA appearing 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

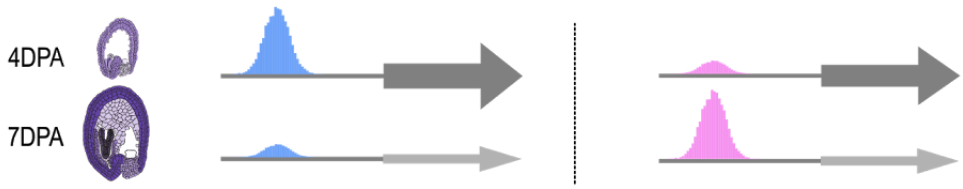
gene set	total	neighboring dDHS expected		neighboring dDHS observed	fold enrichment
Known	48	8	<<	20	2.50*
△Expr, 4DPA vs 7DPA					
Belmonte et al. 2013	4115	685	<<	1120	1.63**
Dean et al. 2011	3423	570	<<	913	1.60**

B



C

genes with greater expression at 4DPA	total	neighboring dDHS more accessible in 4DPA than 7DPA expected		neighboring dDHS more accessible in 4DPA than 7DPA observed	fold enrichment	neighboring dDHS more accessible in 7DPA than 4DPA expected		neighboring dDHS more accessible in 7DPA than 4DPA observed	fold enrichment
Belmonte et al. 2013	1984	122	<<	289	2.37**	225	>>	182	0.81*
Dean et al. 2011	1391	85	<<	166	1.94**	158	>	135	0.85



genes with greater expression at 7DPA	total	neighboring dDHS more accessible in 4DPA than 7DPA expected		neighboring dDHS more accessible in 4DPA than 7DPA observed	fold enrichment	neighboring dDHS more accessible in 7DPA than 4DPA expected		neighboring dDHS more accessible in 7DPA than 4DPA observed	fold enrichment
Belmonte et al. 2013	2131	242	>	126	0.52	130	<<	586	4.48**
Dean et al. 2011	2032	230	>	138	0.60	125	<<	524	4.20**

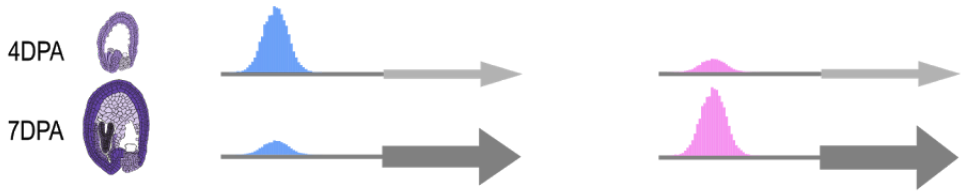
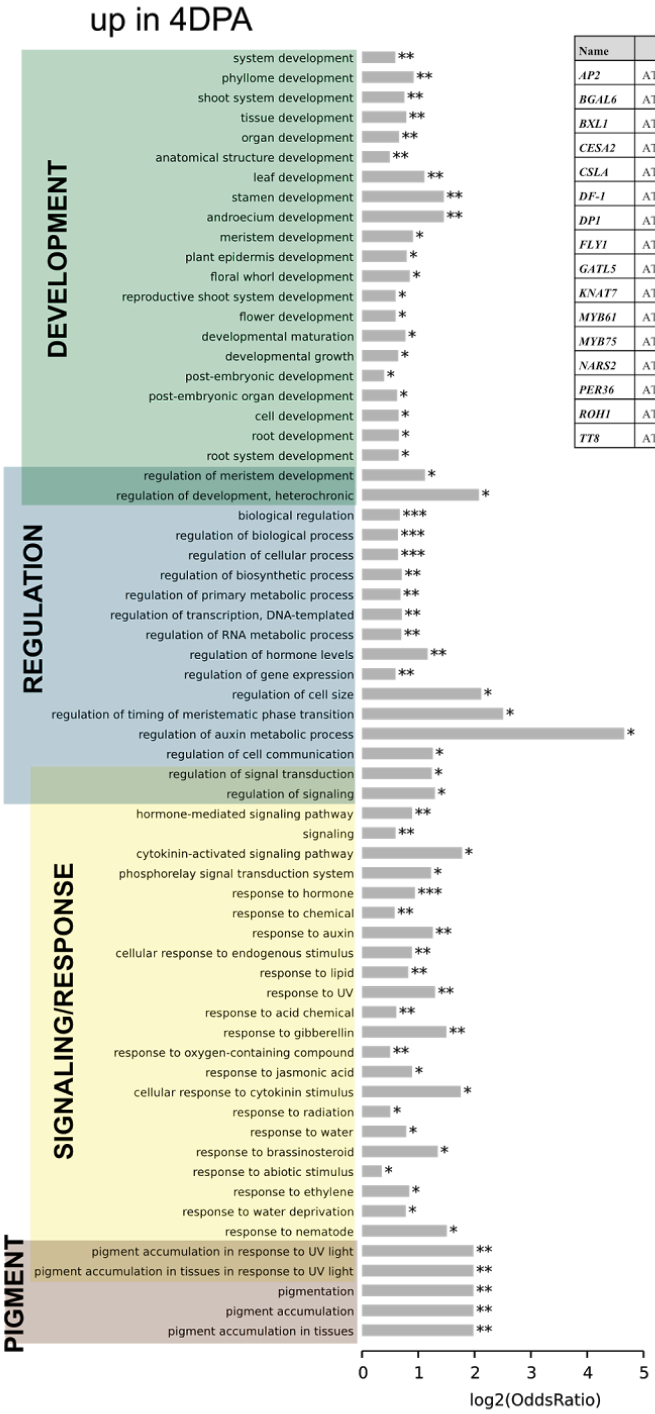


Figure 2. Genes neighboring developmentally dynamic DHSs are often differentially expressed. A, Overlap between the set of genes neighboring dDHSs and 48 known seed coat development genes and genes differentially expressed in seed coat (G. Dean et al. 2011; Belmonte et al. 2013). **B,** Overlap of all four sets of genes. **C,** Genes that are more highly expressed tend to be near more accessible DHSs and vice verse. One asterisk (*) indicates p-value < 0.01. Two asterisks (**) indicate p-value < 10^{-20}.

A



B

Of the 48 Genes known to be involved in seed coat development, 20 neighbor differentially accessible DHSs (4DPA vs 7DPA)

Name	ATG	Gene family	Function	Citation
AP2	AT4G36920	AP2/ERF TF	ovule integument	(Ohto et al., 2009)
BGAL6	AT5G63800	beta-galactosidase 6	pectin modification	(Dean et al., 2007)
BXL1	AT5G49360	B-d-xylosidase	pectin modification	(Arsovski et al., 2009)
CESA2	AT4G39350	cellulose synthase	secondary cell wall cellulose	(Mendu et al., 2011)
CSLA	AT5G22740	cellulose synthase like	mucilage production	(Yu et al., 2014)
DF-1	AT1G76880	putative MYB	mucilage extrusion	(Vasilevski et al., 2012)
DP1	AT4G11180	dirigent protein	lignan synthesis	(Esfandiari et al., 2013)
FLY1	AT4G28370	RING E3 ubiquitin ligase	mucilage modification	(Voiniciuc et al., 2013)
GATL5	AT1G02720	glycosyltransferase	mucilage production	(Kong et al., 2013)
KNAT7	AT1G70510	Knox TF	mucilage, cell morphology	(Bhargava et al., 2013)
MYB61	AT1G09540	MYB TF	mucilage, extrusion	(Penfield et al., 2001)
MYB75	AT1G56650	MYB TF	mucilage, cell morphology	(Bhargava et al., 2013)
NARS2	AT1G52880	NAM/ATAF/NAC TFs	differentiation, cell wall	(Kunieda et al., 2008)
PER36	AT3G50990	class III peroxidase	cell wall, mucilage release	(Kunieda et al., 2013)
ROH1	AT1G63930	DUF793 domain	secretion	(Kulich et al., 2010)
TT8	AT4G09820	bHLH TF	pigmentation, differentiation	(Gonzalez et al., 2009)

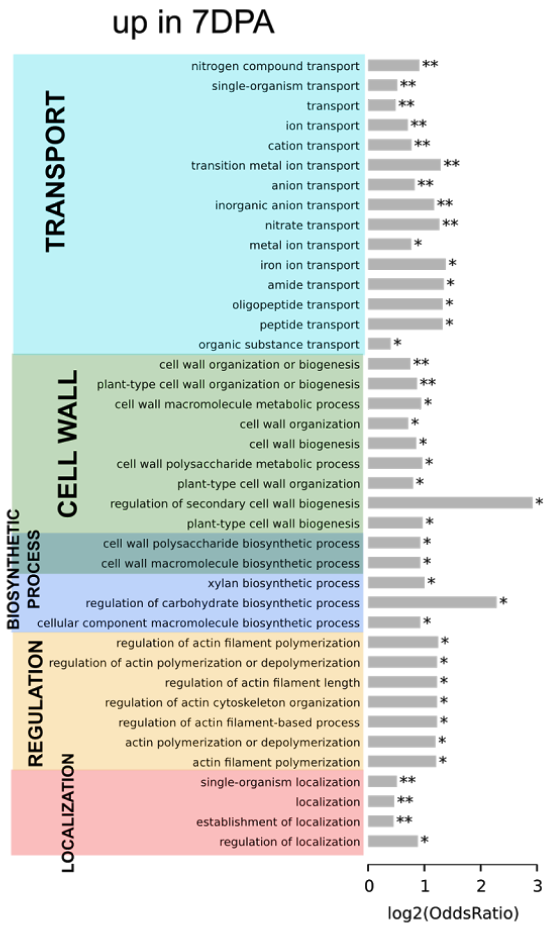


Figure 3. Term enrichment for genes nearest to dynamic DHSs. A, Term enrichment for genes near DHSs that are less (*left*) and more (*right*) accessible at the 7DPA time point. **B,** The twenty genes of the 48 known seed coat development genes that neighbor one or more dynamic DHS.

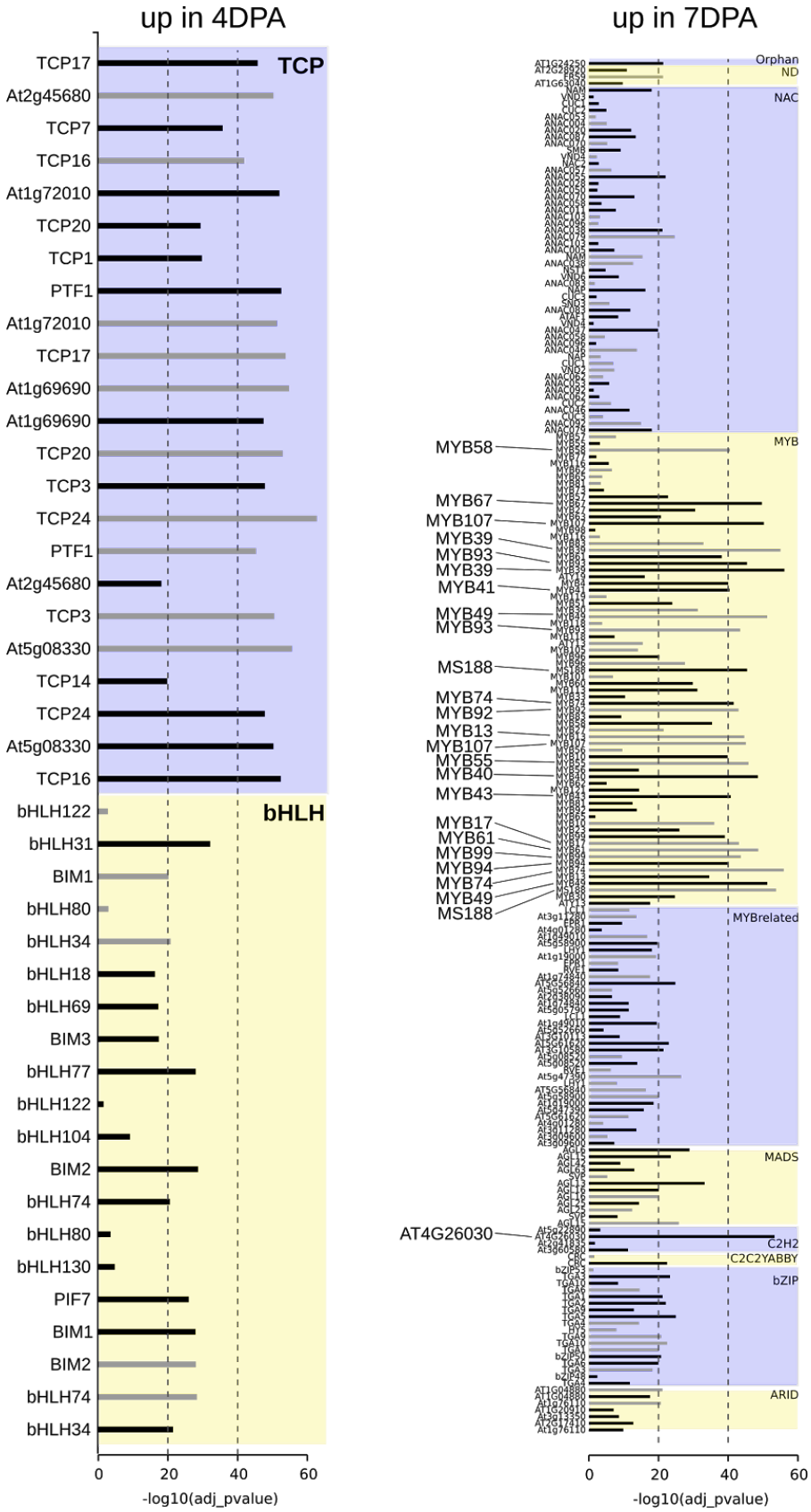

up in 4DPA
TCP
bHLH
TCP17
At2g45680
TCP7
TCP16
At1g72010
TCP20
TCP1
PTF1
At1g72010
TCP17
At1g69690
At1g69690
TCP20
TCP3
TCP24
PTF1
At2g45680
TCP3
At5g08330
TCP14
TCP24
At5g08330
TCP16
bHLH122
bHLH31
BIM1
bHLH80
bHLH34
bHLH18
bHLH69
BIM3
bHLH77
bHLH122
bHLH104
BIM2
bHLH74
bHLH80
bHLH130
PIF7
BIM1
BIM2
bHLH74
bHLH34
0
20
40
60
-log10(adj_pvalue)
up in 7DPA
Orphan
ND
NAC
MYB
MYB58
MYB67
MYB107
MYB39
MYB93
MYB39
MYB41
MYB49
MYB93
MS188
MYB74
MYB92
MYB13
MYB107
MYB55
MYB40
MYB43
MYB17
MYB61
MYB99
MYB94
MYB74
MYB49
MS188
MYBrelated
MADS
AT4G26030
C2H2
C2C2YABBY
bZIP
ARID
0
20
40
60
-log10(adj_pvalue)

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 -log10(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.

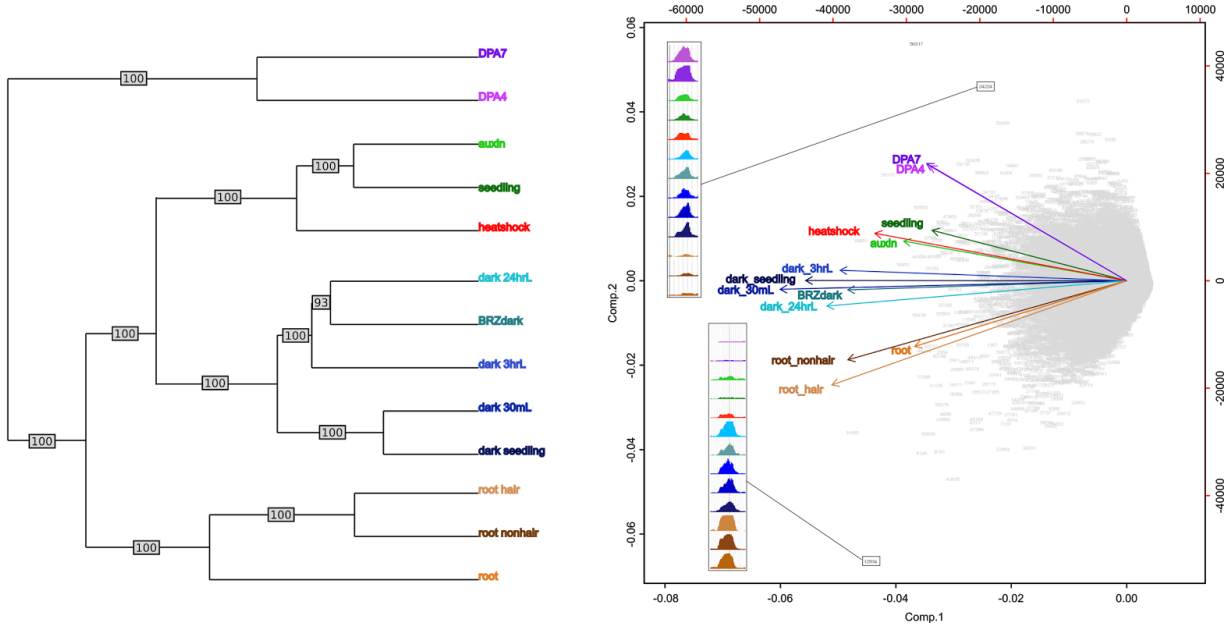


Figure 5. Comparative analysis of diverse DHS landscapes. A, Dendrogram of thirteen samples using accessibility data. **B,** Biplot of Principal Component Analysis of 62,729-DHS by 13-sample matrix. Numbers in gray represent union DHSs (rows in the matrix). The accessibility of all 13 samples are displayed for two uDHSs that were highly informative for distinguishing samples (i.e. markedly differentially accessible).

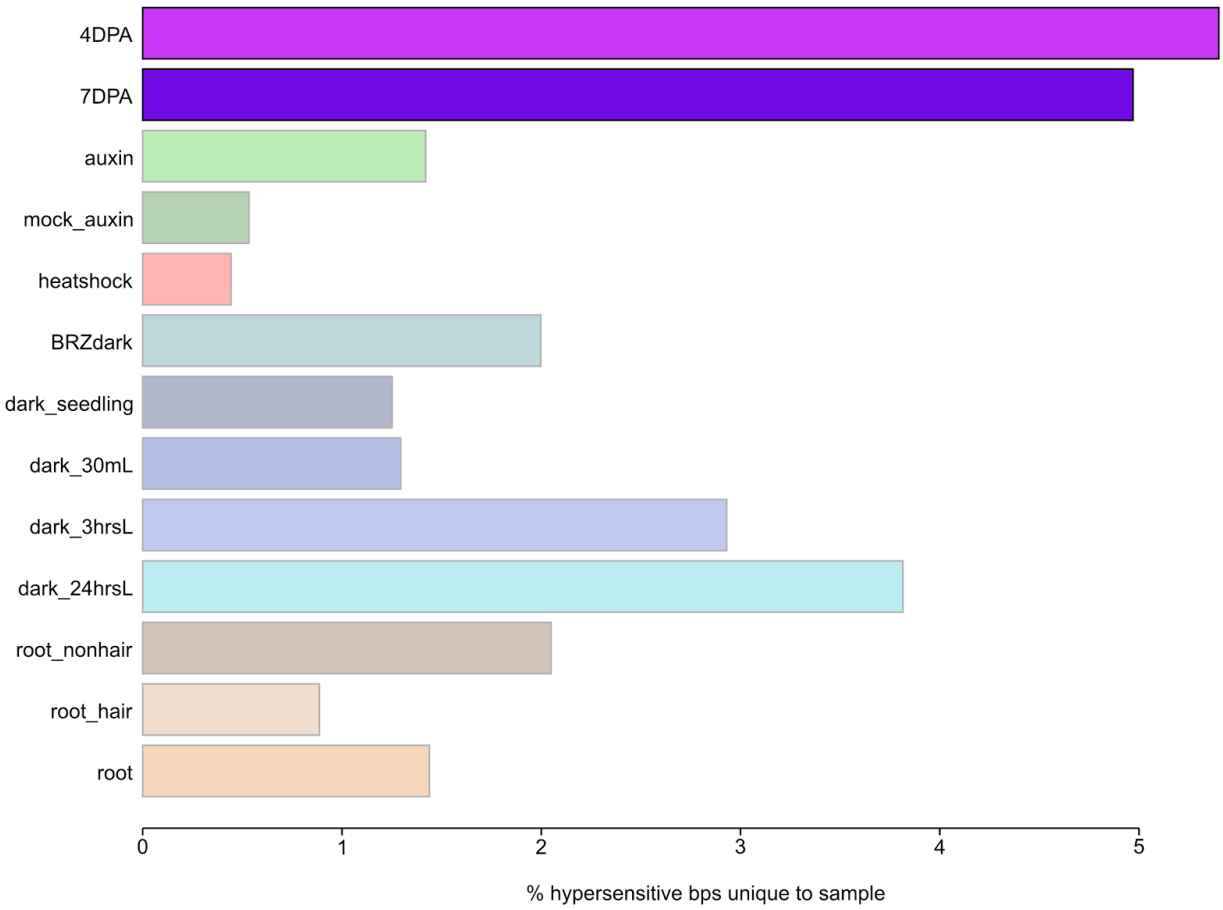


Figure 6. Silique-derived samples contribute the most novel hypersensitive bases of the thirteen samples examined. Shown are the fraction of bps in union DHSs lost if a given sample is excluded from the union set.

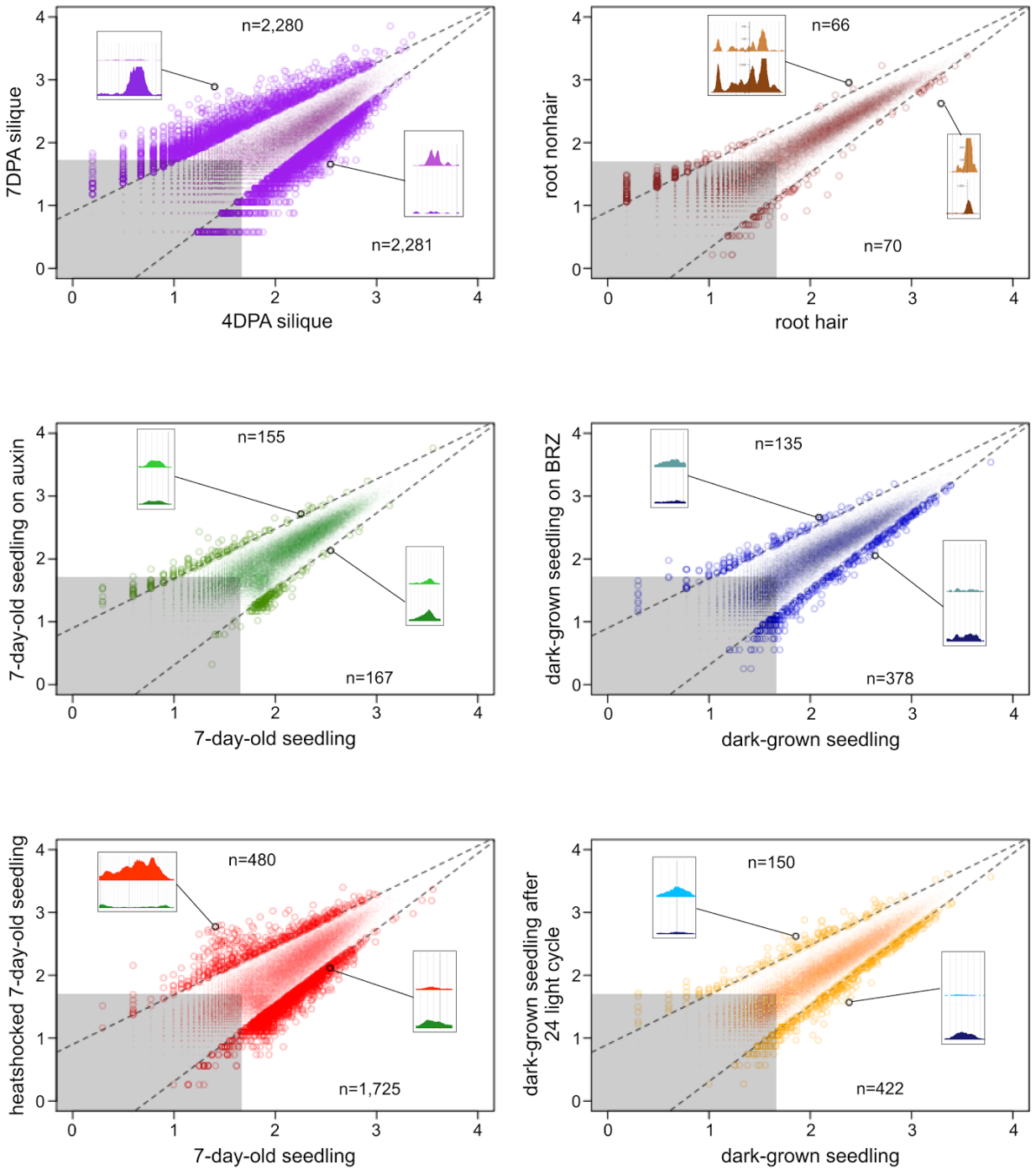

n=2,280
n=2,281
7DPA silique
4DPA silique
n=66
n=70
root nonhair
root hair
n=155
n=167
7-day-old seedling on auxin
7-day-old seedling
n=135
n=378
dark-grown seedling on BRZ
dark-grown seedling
n=480
n=1,725
heatshocked 7-day-old seedling
7-day-old seedling
n=150
n=422
dark-grown seedling after 24 light cycle
dark-grown seedling

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 [*log10(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

Aceituno, Felipe F., Nick Moseyko, Seung Y. Rhee, and Rodrigo A. Gutiérrez. 2008. "The Rules of Gene Expression in Plants: Organ Identity and Gene Body Methylation Are Key Factors for Regulation of Gene Expression in Arabidopsis Thaliana." *BMC Genomics* 9 (September):438.

Alexandre, C. M., J. R. Urton, K. Jean-Baptiste, J. L. Huddleston, M. W. Dorrity, J. T. Cuperus, A. M. Sullivan, et al. 2017. "Complex Relationships between Chromatin Accessibility, Sequence Divergence, and Gene Expression in A. Thaliana." *Molecular Biology and Evolution*, December. https://doi.org/10.1093/molbev/msx326.

Alonso, José M., Anna N. Stepanova, Thomas J. Leisse, Christopher J. Kim, Huaming Chen, Paul Shinn, Denise K. Stevenson, et al. 2003. "Genome-Wide Insertional Mutagenesis of Arabidopsis Thaliana." *Science* 301 (5633):653–57.

Ambawat, Supriya, Poonam Sharma, Neelam R. Yadav, and Ram C. Yadav. 2013. "MYB Transcription Factor Genes as Regulators for Plant Responses: An Overview." *Physiology and Molecular Biology of Plants: An International Journal of Functional Plant Biology* 19 (3):307–21.

Arsovski, Andrej A., Theodore M. Popma, George W. Haughn, Nicholas C. Carpita, Maureen C. McCann, and Tamara L. Western. 2009. "AtBXL1 Encodes a Bifunctional Beta-D-Xylosidase/alpha-L-Arabinofuranosidase Required for Pectic Arabinan Modification in Arabidopsis Mucilage Secretory Cells." *Plant Physiology* 150 (3):1219–34.

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):729–40.

Baniahmad, Aria, Christof Steiner, Anja Carola Köhne, and Rainer Renkawitz. 1990. "Modular Structure of a Chicken Lysozyme Silencer: Involvement of an Unusual Thyroid Hormone Receptor Binding Site." *Cell* 61 (3):505–14.

Belmonte, Mark F., Ryan C. Kirkbride, Sandra L. Stone, Julie M. Pelletier, Anhthu Q. Bui, Edward C. Yeung, Meryl Hashimoto, et al. 2013. "Comprehensive Developmental Profiles of Gene Activity in Regions and Subregions of the Arabidopsis Seed." *Proceedings of the National Academy of Sciences of the United States of America* 110 (5):E435–44.

Ben-Tov, Daniela, Yael Abraham, Shira Stav, Kevin Thompson, Ann Loraine, Rivka Elbaum, Amancio de Souza, Markus Pauly, Joseph J. Kieber, and Smadar Harpaz-Saad. 2015. "COBRA-LIKE2, a Member of the Glycosylphosphatidylinositol-Anchored COBRA-LIKE Family, Plays a Role in Cellulose Deposition in Arabidopsis Seed Coat Mucilage Secretory Cells." *Plant Physiology* 167 (3):711–24.

Bhargava, Apurva, Abdul Ahad, Shucai Wang, Shawn D. Mansfield, George W. Haughn, Carl J. Douglas, and Brian E. Ellis. 2013. "The Interacting MYB75 and KNAT7 Transcription Factors Modulate Secondary Cell Wall Deposition Both in Stems and Seed Coat in Arabidopsis." *Planta* 237 (5):1199–1211.

Burn, Joanne E., Ursula A. Hurley, Rosemary J. Birch, Tony Arioli, Ann Cork, and Richard E. Williamson. 2002. "The Cellulose-Deficient Arabidopsis Mutant rsw3 Is Defective in a Gene Encoding a Putative Glucosidase II, an Enzyme Processing N-Glycans during ER Quality Control." *The Plant Journal: For Cell and Molecular Biology* 32 (6):949–60.

Caffall, Kerry H., Sivakumar Pattathil, Sarah E. Phillips, Michael G. Hahn, and Debra Mohnen. 2009. "Arabidopsis Thaliana T-DNA Mutants Implicate GAUT Genes in the Biosynthesis of Pectin and Xylan in Cell Walls and Seed Testa." *Molecular Plant* 2 (5):1000–1014.

Carlson, Marc. 2016. *org.At.tair.db: Genome Wide Annotation for Arabidopsis*. http://bioconductor.org/packages/release/data/annotation/html/org.At.tair.db.html.

Chung, Jay H., Adam C. Bell, and Gary Felsenfeld. 1997. "Characterization of the Chicken β-Globin Insulator." *Proceedings of the National Academy of Sciences* 94 (2):575–80.

Clark, Richard M., Tina Nussbaum Wagler, Pablo Quijada, and John Doebley. 2006. "A Distant Upstream Enhancer at the Maize Domestication Gene tb1 Has Pleiotropic Effects on Plant and Inflorescent Architecture." *Nature Genetics* 38 (5):594–97.

Copenhaver, G. P., and C. S. Pikaard. 1996. "RFLP and Physical Mapping with an rDNA-Specific Endonuclease Reveals That Nucleolus Organizer Regions of Arabidopsis Thaliana Adjoin the Telomeres on Chromosomes 2 and 4." *The Plant Journal: For Cell and Molecular Biology* 9 (2):259–72.

Cubas, P., N. Lauter, J. Doebley, and E. Coen. 1999. "The TCP Domain: A Motif Found in Proteins Regulating Plant Growth and Development." *The Plant Journal: For Cell and Molecular Biology* 18 (2):215–22.

Daugherty, Aaron C., Robin Yeo, Jason D. Buenrostro, and William J. Greenleaf. 2017. "Chromatin Accessibility Dynamics Reveal Novel Functional Enhancers in C. Elegans." *bioRxiv*, June. https://doi.org/10.1101/088732.

Deal, Roger B., and Steven Henikoff. 2010. "A Simple Method for Gene Expression and Chromatin Profiling of Individual Cell Types within a Tissue." *Developmental Cell* 18 (6):1030–40.

Dean, Gillian, Yongguo Cao, Daoquan Xiang, Nicholas J. Provart, Larissa Ramsay, Abdul Ahad, Rick White, Gopalan Selvaraj, Raju Datla, and George Haughn. 2011. "Analysis of Gene Expression Patterns during Seed Coat Development in Arabidopsis." *Molecular Plant* 4 (6):1074–91.

Dean, Gillian H., Huanquan Zheng, Jagdish Tewari, Jun Huang, Diana S. Young, Yeen Ting Hwang, Tamara L. Western, et al. 2007. "The Arabidopsis MUM2 Gene Encodes a Beta-Galactosidase Required for the Production of Seed Coat Mucilage with Correct Hydration Properties." *The Plant Cell* 19 (12):4007–21.

Elgin, S. C. 1988. "The Formation and Function of DNase I Hypersensitive Sites in the Process of Gene Activation." *The Journal of Biological Chemistry* 263 (36). American Society for Biochemistry and Molecular Biology:19259–62.

Esfandiari, Elahe, Zhaoqing Jin, Ashraf Abdeen, Jonathan S. Griffiths, Tamara L. Western, and George W. Haughn. 2013. "Identification and Analysis of an Outer-Seed-Coat-Specific

Promoter from Arabidopsis Thaliana." *Plant Molecular Biology* 81 (1-2):93–104.
Falcon, S., and R. Gentleman. 2007. "Using GOstats to Test Gene Lists for GO Term Association." *Bioinformatics* 23 (2):257–58.
Francoz, Edith, Philippe Ranocha, Vincent Burlat, and Christophe Dunand. 2015. "Arabidopsis Seed Mucilage Secretory Cells: Regulation and Dynamics." *Trends in Plant Science* 20 (8):515–24.
Galway, M. E., J. D. Masucci, A. M. Lloyd, V. Walbot, R. W. Davis, and J. W. Schiefelbein. 1994. "The TTG Gene Is Required to Specify Epidermal Cell Fate and Cell Patterning in the Arabidopsis Root." *Developmental Biology* 166 (2):740–54.
García-Fayos, Patricio, Esther Bochet, and Artemi Cerdà. 2010. "Seed Removal Susceptibility through Soil Erosion Shapes Vegetation Composition." *Plant and Soil* 334 (1-2). Springer Netherlands:289–97.
Garwood, Nancy C. 1985. "The Role of Mucilage in the Germination of Cuipo, Cavanillesia Platanifolia (H. & B.) H. B. K. (Bombacaceae), a Tropical Tree." *American Journal of Botany* 72 (7). Botanical Society of America:1095–1105.
Gendre, Delphine, Heather E. McFarlane, Errin Johnson, Gregory Mouille, Andreas Sjödin, Jaesung Oh, Gabriel Levesque-Tremblay, Yoichiro Watanabe, Lacey Samuels, and Rishikesh P. Bhalerao. 2013. "Trans-Golgi Network Localized ECHIDNA/Ypt Interacting Protein Complex Is Required for the Secretion of Cell Wall Polysaccharides in Arabidopsis." *The Plant Cell* 25 (7):2633–46.
Gonzalez, Antonio, John Mendenhall, Yujia Huo, and Alan Lloyd. 2009. "TTG1 Complex MYBs, MYB5 and TT2, Control Outer Seed Coat Differentiation." *Developmental Biology* 325 (2):412–21.
Groudine, M., and H. Weintraub. 1982. "Propagation of Globin DNAase I-Hypersensitive Sites in Absence of Factors Required for Induction: A Possible Mechanism for Determination." *Cell* 30 (1):131–39.
Gutterman, Yitzchak, and Shachar Shem-Tov. 1997. "THE EFFICIENCY OF THE STRATEGY OF MUCILAGINOUS SEEDS OF SOME COMMON ANNUALS OF THE NEGEV ADHERING TO THE SOIL CRUST TO DELAY COLLECTION BY ANTS." *Israel Journal of Plant Sciences* 45 (4). Taylor & Francis:317–27.
Harpaz-Saad, Smadar, Heather E. McFarlane, Shouling Xu, Uday K. Divi, Bronwen Forward, Tamara L. Western, and Joseph J. Kieber. 2011. "Cellulose Synthesis via the FEI2 RLK/SOS5 Pathway and Cellulose Synthase 5 Is Required for the Structure of Seed Coat Mucilage in Arabidopsis." *The Plant Journal: For Cell and Molecular Biology* 68 (6):941–53.
Harpaz-Saad, Smadar, Tamara L. Western, and Joseph J. Kieber. 2012. "The FEI2-SOS5 Pathway and CELLULOSE SYNTHASE 5 Are Required for Cellulose Biosynthesis in the Arabidopsis Seed Coat and Affect Pectin Mucilage Structure." *Plant Signaling & Behavior* 7 (2):285–88.
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.

Honma, T., and K. Goto. 2001. "Complexes of MADS-Box Proteins Are Sufficient to Convert Leaves into Floral Organs." *Nature* 409 (6819):525–29.

Huang, Jun, Danisha DeBowles, Elahe Esfandiari, Gillian Dean, Nicholas C. Carpita, and George W. Haughn. 2011. "The Arabidopsis Transcription Factor LUH/MUM1 Is Required for Extrusion of Seed Coat Mucilage." *Plant Physiology* 156 (2):491–502.

John, Sam, Peter J. Sabo, Theresa K. Canfield, Kristen Lee, Shinny Vong, Molly Weaver, Hao Wang, et al. 2013. "Genome-Scale Mapping of DNase I Hypersensitivity." *Current Protocols in Molecular Biology / Edited by Frederick M. Ausubel ... [et Al.]* Chapter 27 (July):Unit 21.27.

John, Sam, Peter J. Sabo, Robert E. Thurman, Myong-Hee Sung, Simon C. Biddie, Thomas A. Johnson, Gordon L. Hager, and John A. Stamatoyannopoulos. 2011a. "Chromatin Accessibility Pre-Determines Glucocorticoid Receptor Binding Patterns." *Nature Genetics* 43 (3):264–68.

———. 2011b. "Chromatin Accessibility Pre-Determines Glucocorticoid Receptor Binding Patterns." *Nature Genetics* 43 (3):264–68.

Johnson, Cameron S., Ben Kolevski, and David R. Smyth. 2002. "TRANSPARENT TESTA GLABRA2, a Trichome and Seed Coat Development Gene of Arabidopsis, Encodes a WRKY Transcription Factor." *The Plant Cell* 14 (6):1359–75.

Karssen, C. M., D. L. Brinkhorst-van der Swan, A. E. Breekland, and M. Koornneef. 1983. "Induction of Dormancy during Seed Development by Endogenous Abscisic Acid: Studies on Abscisic Acid Deficient Genotypes of Arabidopsis Thaliana (L.) Heynh." *Planta* 157 (2):158–65.

Kim, Young-Cheon, Masatoshi Nakajima, Akira Nakayama, and Isomaro Yamaguchi. 2005. "Contribution of Gibberellins to the Formation of Arabidopsis Seed Coat through Starch Degradation." *Plant & Cell Physiology* 46 (8):1317–25.

Kong, Yingzhen, Gongke Zhou, Ashraf A. Abdeen, James Schafhauser, Beth Richardson, Melani A. Atmodjo, Jiyoung Jung, et al. 2013. "GALACTURONOSYLTRANSFERASE-LIKE5 Is Involved in the Production of Arabidopsis Seed Coat Mucilage." *Plant Physiology* 163 (3):1203–17.

Kulich, Ivan, Rex Cole, Edita Drdová, Fatima Cvrcková, Ales Soukup, John Fowler, and Viktor Zárský. 2010. "Arabidopsis Exocyst Subunits SEC8 and EXO70A1 and Exocyst Interactor ROH1 Are Involved in the Localized Deposition of Seed Coat Pectin." *The New Phytologist* 188 (2):615–25.

Kunieda, Tadashi, Nobutaka Mitsuda, Masaru Ohme-Takagi, Seiji Takeda, Mitsuhiro Aida, Masao Tasaka, Maki Kondo, Mikio Nishimura, and Ikuko Hara-Nishimura. 2008. "NAC Family Proteins NARS1/NAC2 and NARS2/NAM in the Outer Integument Regulate Embryogenesis in Arabidopsis." *The Plant Cell* 20 (10):2631–42.

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

Ikuko Hara-Nishimura. 2013. “Spatiotemporal Secretion of PEROXIDASE36 Is Required for Seed Coat Mucilage Extrusion in Arabidopsis.” *The Plant Cell* 25 (4):1355–67.

Le, Brandon H., Chen Cheng, Anhthu Q. Bui, Javier A. Wagmaister, Kelli F. Henry, Julie Pelletier, Linda Kwong, et al. 2010. “Global Analysis of Gene Activity during Arabidopsis Seed Development and Identification of Seed-Specific Transcription Factors.” *Proceedings of the National Academy of Sciences of the United States of America* 107 (18):8063–70.

Leon-Kloosterziel, K. M., C. J. Keijzer, and M. Koornneef. 1994. “A Seed Shape Mutant of Arabidopsis That Is Affected in Integument Development.” *The Plant Cell* 6 (3):385–92.

Lin, X., S. Kaul, S. Rounsley, T. P. Shea, M. I. Benito, C. D. Town, C. Y. Fujii, et al. 1999. “Sequence and Analysis of Chromosome 2 of the Plant Arabidopsis Thaliana.” *Nature* 402 (6763):761–68.

Li, Song Feng, Olga Nicolaou Milliken, Hanh Pham, Reg Seyit, Ross Napoli, Jeremy Preston, Anna M. Koltunow, and Roger W. Parish. 2009. “The Arabidopsis MYB5 Transcription Factor Regulates Mucilage Synthesis, Seed Coat Development, and Trichome Morphogenesis.” *The Plant Cell* 21 (1):72–89.

Love, Michael I., Wolfgang Huber, and Simon Anders. 2014. “Moderated Estimation of Fold Change and Dispersion for RNA-Seq Data with DESeq2.” *Genome Biology* 15 (12):550.

Martín-Trillo, Mar, and Pilar Cubas. 2010. “TCP Genes: A Family Snapshot Ten Years Later.” *Trends in Plant Science* 15 (1):31–39.

McFarlane, Heather E., Robin E. Young, Geoffrey O. Wasteneys, and A. Lacey Samuels. 2008. “Cortical Microtubules Mark the Mucilage Secretion Domain of the Plasma Membrane in Arabidopsis Seed Coat Cells.” *Planta* 227 (6):1363–75.

McLeay, Robert C., and Timothy L. Bailey. 2010. “Motif Enrichment Analysis: A Unified Framework and an Evaluation on ChIP Data.” *BMC Bioinformatics* 11 (1):165.

Mendu, Venugopal, Jonathan S. Griffiths, Staffan Persson, Jozsef Stork, A. Bruce Downie, Cătălin Voiniciuc, George W. Haughn, and Seth DeBolt. 2011. “Subfunctionalization of Cellulose Synthases in Seed Coat Epidermal Cells Mediates Secondary Radial Wall Synthesis and Mucilage Attachment.” *Plant Physiology* 157 (1). American Society of Plant Biologists:441–53.

Moffatt, Barbara A., Yvonne Y. Stevens, Michael S. Allen, Jamie D. Snider, Luiz A. Pereira, Margarita I. Todorova, Peter S. Summers, Elizabeth A. Weretilnyk, Luke Martin-McCaffrey, and Conrad Wagner. 2002. “Adenosine Kinase Deficiency Is Associated with Developmental Abnormalities and Reduced Transmethylation.” *Plant Physiology* 128 (3):812–21.

Morgan, W. D., G. T. Williams, R. I. Morimoto, J. Greene, R. E. Kingston, and R. Tjian. 1987. “Two Transcriptional Activators, CCAAT-Box-Binding Transcription Factor and Heat Shock Transcription Factor, Interact with a Human hsp70 Gene Promoter.” *Molecular and Cellular Biology* 7 (3):1129–38.

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

Molecular Evolution 65 (1):23–33.
Neph, Shane, M. Scott Kuehn, Alex P. Reynolds, Eric Haugen, Robert E. Thurman, Audra K. Johnson, Eric Rynes, et al. 2012. "BEDOPS: High-Performance Genomic Feature Operations." *Bioinformatics* 28 (14):1919–20.
North, Helen M., Adeline Berger, Susana Saez-Aguayo, and Marie-Christine Ralet. 2014. "Understanding Polysaccharide Production and Properties Using Seed Coat Mutants: Future Perspectives for the Exploitation of Natural Variants." *Annals of Botany* 114 (6). Oxford University Press:1251–63.
Ohto, Masa-Aki, Sandra K. Floyd, Robert L. Fischer, Robert B. Goldberg, and John J. Harada. 2009. "Effects of APETALA2 on Embryo, Endosperm, and Seed Coat Development Determine Seed Size in Arabidopsis." *Sexual Plant Reproduction* 22 (4):277–89.
O'Malley, Ronan C., Shao-Shan Carol Huang, Liang Song, Mathew G. Lewsey, Anna Bartlett, Joseph R. Nery, Mary Galli, Andrea Gallavotti, and Joseph R. Ecker. 2016. "Cistrome and Epicistrome Features Shape the Regulatory DNA Landscape." *Cell* 165 (5):1280–92.
Paradis, Emmanuel, Julien Claude, and Korbinian Strimmer. 2004. "APE: Analyses of Phylogenetics and Evolution in R Language." *Bioinformatics* 20 (2):289–90.
Pastore, Jennifer J., Andrea Limpuangthip, Nobutoshi Yamaguchi, Miin-Feng Wu, Yi Sang, Soon-Ki Han, Lauren Malaspina, Natasha Chavdaroff, Ayako Yamaguchi, and Doris Wagner. 2011. "LATE MERISTEM IDENTITY2 Acts Together with LEAFY to Activate APETALA1." *Development* 138 (15):3189–98.
Penfield, S., R. C. Meissner, D. A. Shoue, N. C. Carpita, and M. W. Bevan. 2001. "MYB61 Is Required for Mucilage Deposition and Extrusion in the Arabidopsis Seed Coat." *The Plant Cell* 13 (12):2777–91.
Pinyopich, Anusak, Gary S. Ditta, Beth Savidge, Sarah J. Liljegren, Elvira Baumann, Ellen Wisman, and Martin F. Yanofsky. 2003. "Assessing the Redundancy of MADS-Box Genes during Carpel and Ovule Development." *Nature* 424 (6944). Nature Publishing Group:85–88.
Rani, Sapa Hima, T. H. Anantha Krishna, Saikat Saha, Arvind Singh Negi, and Ram Rajasekharan. 2010. "Defective in Cuticular Ridges (DCR) of Arabidopsis Thaliana, a Gene Associated with Surface Cutin Formation, Encodes a Soluble Diacylglycerol Acyltransferase." *The Journal of Biological Chemistry* 285 (49):38337–47.
Rautengarten, Carsten, Björn Usadel, Lutz Neumetzler, Jürgen Hartmann, Dirk Büssis, and Thomas Altmann. 2008. "A Subtilisin-like Serine Protease Essential for Mucilage Release from Arabidopsis Seed Coats." *The Plant Journal: For Cell and Molecular Biology* 54 (3):466–80.
Saez-Aguayo, Susana, Marie-Christine Ralet, Adeline Berger, Lucy Botran, David Ropartz, Annie Marion-Poll, and Helen M. North. 2013. "PECTIN METHYLESTERASE INHIBITOR6 Promotes Arabidopsis Mucilage Release by Limiting Methylesterification of Homogalacturonan in Seed Coat Epidermal Cells." *The Plant Cell* 25 (1):308–23.
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.

Stergachis, Andrew B., Shane Neph, Alex Reynolds, Richard Humbert, Brady Miller, Sharon L. Paige, Benjamin Vernot, et al. 2013. "Developmental Fate and Cellular Maturity Encoded in Human Regulatory DNA Landscapes." *Cell* 154 (4):888–903.

Stork, Jozsef, Darby Harris, Jonathan Griffiths, Brian Williams, Fred Beisson, Yonghua Li-Beisson, Venugopal Mendu, George Haughn, and Seth DeBolt. 2010. "CELLULOSE SYNTHASE9 Serves a Nonredundant Role in Secondary Cell Wall Synthesis in Arabidopsis Epidermal Testa Cells." *Plant Physiology* 153 (2). American Society of Plant Biologists:580–89.

Stracke, R., M. Werber, and B. Weisshaar. 2001. "The R2R3-MYB Gene Family in Arabidopsis Thaliana." *Current Opinion in Plant Biology* 4 (5):447–56.

Sullivan, Alessandra M., Andrej A. Arsovski, Janne Lempe, Kerry L. Bubb, Matthew T. Weirauch, Peter J. Sabo, Richard Sandstrom, et al. 2014. "Mapping and Dynamics of Regulatory DNA and Transcription Factor Networks in A. Thaliana." *Cell Reports* 8 (6):2015–30.

Sullivan, Stuart, Marie-Christine Ralet, Adeline Berger, Eugene Diatloff, Volker Bischoff, Martine Gonneau, Annie Marion-Poll, and Helen M. North. 2011. "CESA5 Is Required for the Synthesis of Cellulose with a Role in Structuring the Adherent Mucilage of Arabidopsis Seeds." *Plant Physiology* 156 (4):1725–39.

Talbot, D., P. Collis, M. Antoniou, M. Vidal, F. Grosveld, and D. R. Greaves. 1989. "A Dominant Control Region from the Human Beta-Globin Locus Conferring Integration Site-Independent Gene Expression." *Nature* 338 (6213):352–55.

Thomas, Sean, Xiao-Yong Li, Peter J. Sabo, Richard Sandstrom, Robert E. Thurman, Theresa K. Canfield, Erika Giste, et al. 2011. "Dynamic Reprogramming of Chromatin Accessibility during Drosophilaembryo Development." *Genome Biology* 12 (5):R43.

Thurman, Robert E., Eric Rynes, Richard Humbert, Jeff Vierstra, Matthew T. Maurano, Eric Haugen, Nathan C. Sheffield, et al. 2012. "The Accessible Chromatin Landscape of the Human Genome." *Nature* 489 (7414):75–82.

Tominaga-Wada, Rumi, Mineko Iwata, Junji Sugiyama, Toshihisa Kotake, Tetsuya Ishida, Ryusuke Yokoyama, Kazuhiko Nishitani, Kiyotaka Okada, and Takuji Wada. 2009. "The GLABRA2 Homeodomain Protein Directly Regulates CESA5 and XTH17 Gene Expression in Arabidopsis Roots." *The Plant Journal: For Cell and Molecular Biology* 60 (3):564–74.

Usadel, Björn, Anja M. Kuschinsky, Mario G. Rosso, Nora Eckermann, and Markus Pauly. 2004. "RHM2 Is Involved in Mucilage Pectin Synthesis and Is Required for the Development of the Seed Coat in Arabidopsis." *Plant Physiology* 134 (1):286–95.

Vasilevski, Aleksandar, Federico M. Giorgi, Luca Bertinetti, and Björn Usadel. 2012. "LASSO Modeling of the Arabidopsis Thaliana Seed/seedling Transcriptome: A Model Case for Detection of Novel Mucilage and Pectin Metabolism Genes." *Molecular bioSystems* 8

(10):2566–74.

Voiniciuc, Catalin, Gillian H. Dean, Jonathan S. Griffiths, Kerstin Kirchsteiger, Yeen Ting Hwang, Alan Gillett, Graham Dow, Tamara L. Western, Mark Estelle, and George W. Haughn. 2013. "Flying saucer1 Is a Transmembrane RING E3 Ubiquitin Ligase That Regulates the Degree of Pectin Methylesterification in Arabidopsis Seed Mucilage." *The Plant Cell* 25 (3):944–59.

Walker, Murray, Muhammad Tehseen, Monika S. Doblin, Filomena A. Pettolino, Sarah M. Wilson, Antony Bacic, and John F. Golz. 2011. "The Transcriptional Regulator LEUNIG_HOMOLOG Regulates Mucilage Release from the Arabidopsis Testa." *Plant Physiology* 156 (1):46–60.

Wang, Shui, Jia-Wei Wang, Nan Yu, Chun-Hong Li, Bin Luo, Jin-Ying Gou, Ling-Jian Wang, and Xiao-Ya Chen. 2004. "Control of Plant Trichome Development by a Cotton Fiber MYB Gene." *The Plant Cell* 16 (9):2323–34.

Wei, Baoye, Jinzhe Zhang, Changxu Pang, Hao Yu, Dongshu Guo, Hao Jiang, Mingxin Ding, et al. 2015. "The Molecular Mechanism of Sporocyteless/nozzle in Controlling Arabidopsis Ovule Development." *Cell Research* 25 (1):121–34.

Western, Tamara L., Diana S. Young, Gillian H. Dean, Wei Ling Tan, A. Lacey Samuels, and George W. Haughn. 2004. "MUCILAGE-MODIFIED4 Encodes a Putative Pectin Biosynthetic Enzyme Developmentally Regulated by APETALA2, TRANSPARENT TESTA GLABRA1, and GLABRA2 in the Arabidopsis Seed Coat." *Plant Physiology* 134 (1):296–306.

Western, T. L., D. J. Skinner, and G. W. Haughn. 2000. "Differentiation of Mucilage Secretory Cells of the Arabidopsis Seed Coat." *Plant Physiology* 122 (2):345–56.

Windsor, J. B., V. V. Symonds, J. Mendenhall, and A. M. Lloyd. 2000. "Arabidopsis Seed Coat Development: Morphological Differentiation of the Outer Integument." *The Plant Journal: For Cell and Molecular Biology* 22 (6):483–93.

Wu, C., P. M. Bingham, K. J. Livak, R. Holmgren, and S. C. Elgin. 1979. "The Chromatin Structure of Specific Genes: I. Evidence for Higher Order Domains of Defined DNA Sequence." *Cell* 16 (4):797–806.

Wu, C., Y. C. Wong, and S. C. Elgin. 1979. "The Chromatin Structure of Specific Genes: II. Disruption of Chromatin Structure during Gene Activity." *Cell* 16 (4):807–14.

Yang, Xuejun, Ming Dong, and Zhenying Huang. 2010. "Role of Mucilage in the Germination of Artemisia Sphaerocephala (Asteraceae) Achenes Exposed to Osmotic Stress and Salinity." *Plant Physiology and Biochemistry: PPB / Societe Francaise de Physiologie Vegetale* 48 (2-3):131–35.

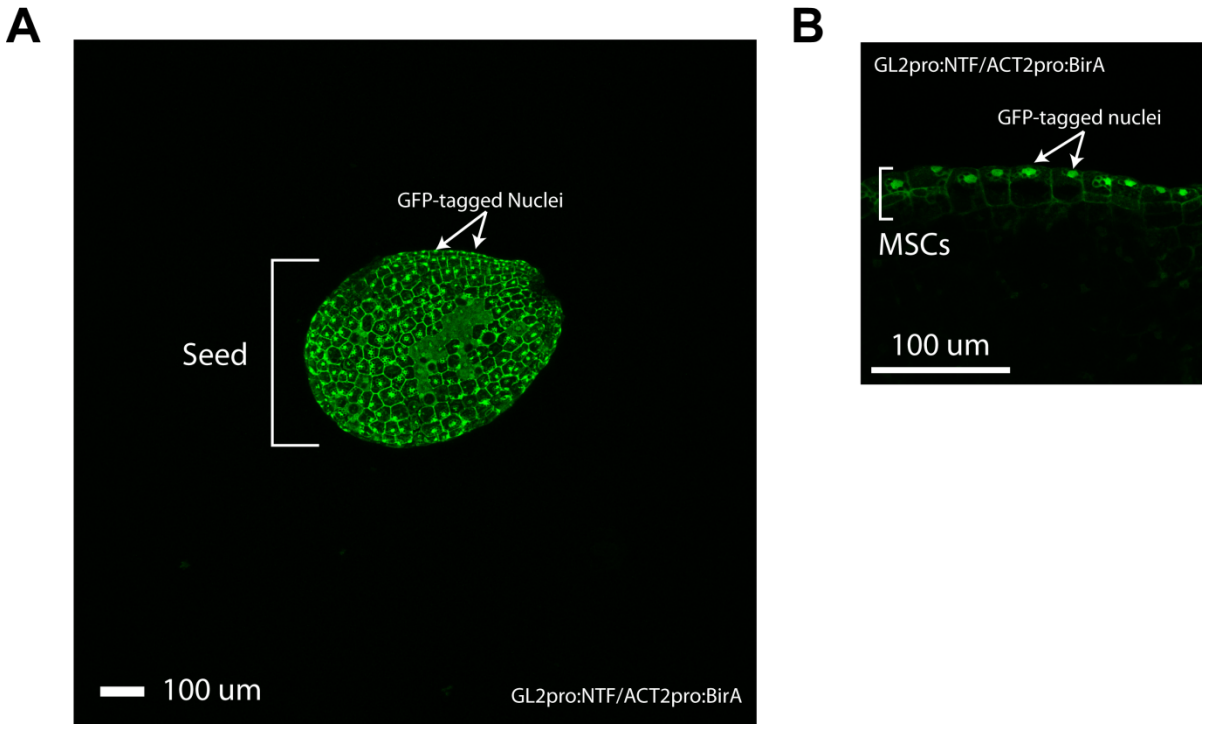
Yang, Xuejun, Wenhao Zhang, Ming Dong, Ivan Boubriak, and Zhenying Huang. 2011. "The Achene Mucilage Hydrated in Desert Dew Assists Seed Cells in Maintaining DNA Integrity: Adaptive Strategy of Desert Plant Artemisia Sphaerocephala." *PloS One* 6 (9). Public Library of Science:e24346.

Yu, Li, Dachuan Shi, Junling Li, Yingzhen Kong, Yanchong Yu, Guohua Chai, Ruibo Hu, Juan

Wang, Michael G. Hahn, and Gongke Zhou. 2014. “CELLULOSE SYNTHASE-LIKE A2, a Glucomannan Synthase, Is Involved in Maintaining Adherent Mucilage Structure in Arabidopsis Seed.” *Plant Physiology* 164 (4):1842–56.

Zabidi, Muhammad A., Cosmas D. Arnold, Katharina Schernhuber, Michaela Pagani, Martina Rath, Olga Frank, and Alexander Stark. 2015. “Enhancer-Core-Promoter Specificity Separates Developmental and Housekeeping Gene Regulation.” *Nature* 518 (7540):556–59.

Supplemental Materials

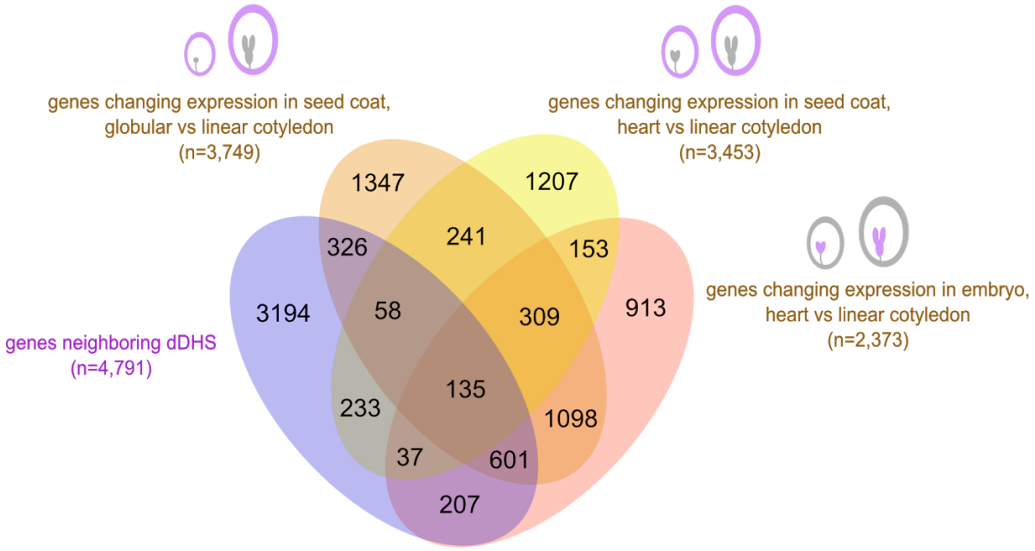


Supplemental Figure 1. Confocal microscopy of INTACT-tagged nuclei in seed coat epidermis. A. Confocal of whole seed at 4DPA from the INTACT line *GL2$_{pro}$:NTF/ACT2$_{pro}$:BirA* (Deal and Henikoff, 2010). GFP-fluorescing nuclei are evident across the seed coat epidermis. Scale is 100um. **B.** Confocal 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.

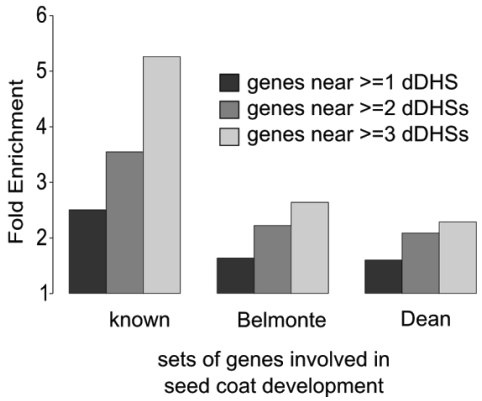
A

gene set	total	neighboring dDHS expected		neighboring dDHS observed	fold enrichment
tissue = seed coat					
globular vs linear cotyledon	4,115	685	<<	1120	1.63**
heart vs linear cotyledon	3,453	575	<<	980	1.58**
tissue = embryo					
heart vs linear cotyledon	2,373	395	<	463	1.17*

B



C



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).