Mapping and Dynamics of Regulatory DNA in Maturing *Arabidopsis thaliana* Siliques

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]*
and Christine Queitsch[1]

[1] Department of Genome Sciences, University of Washington, Seattle, WA, United States, [2] Department of Biology, University of Washington, Seattle, WA, United States

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 functions of epidermal cells have made them a model for connecting transcriptional events to development in a range of model systems. In *Arabidopsis thaliana* and many other plants, fertilization triggers differentiation of specialized epidermal seed coat cells that have a unique morphology caused by large extracellular deposits of polysaccharides. Here, we used DNase I-seq to generate regulatory landscapes of *A. thaliana* seeds at two critical time points in seed coat maturation (4 and 7 DPA), enriching for seed coat cells with the INTACT method. We found over 3,000 developmentally dynamic regulatory DNA elements and explored their relationship with nearby gene expression. The dynamic regulatory elements were enriched for motifs for several transcription factors families; most notably the TCP family at the earlier time point and the MYB family at the later one. To assess the extent to which the observed regulatory sites in seeds added to previously known regulatory sites in *A. thaliana,* we compared our data to 11 other data sets generated with 7-day-old seedlings for diverse tissues and conditions. Surprisingly, over a quarter of the regulatory, i.e. accessible, bases observed in seeds were novel. Notably, plant regulatory landscapes from different tissues, cell types, or developmental stages were more dynamic than those generated from bulk tissue in response to environmental perturbations, highlighting the importance of extending studies of regulatory DNA to single tissues and cell types during development.

Keywords: regulatory DNA, *Arabidopsis thaliana*, seed development, seed coat maturation, open chromatin

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 regulating gene expression. 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 (Wu et al., 1979a; Wu et al.,

1979b; Banerji et al., 1983; Talbot et al., 1989; Baniahmad et al., 1990; Chung et al., 1997; Thurman et al., 2012). Our previous analyses of regulatory DNA and its dynamics in *Arabidopsis thaliana* largely focused on identifying regulatory networks and divergence of regulatory DNA in whole seedlings (Sullivan et al., 2014). Our method, which relies on the INTACT (isolation of nuclei tagged in specific cell types) method of preparing nuclei (Deal and Henikoff, 2010), lends itself to investigating the regulatory landscape of nuclei enriched for certain cell types. Cell-type–enriched, and ideally cell-type–specific, 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 critical developmental time points, 4 and 7 days post-anthesis, enriching for seed coat cells as they transition from the non-mucous-secreting state to the mucous-secreting state.

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 (Western et al., 2000; Windsor et al., 2000). Although the function of mucilage depends on the species and the environmental context (Garwood, 1985; Gutterman and Shem-Tov, 1997; García-Fayos et al., 2010; Yang et al., 2010; Yang et al., 2011), mucilage is generally thought to protect the emerging seedling and facilitate its germination. While we aim to identify the *cis*-regulatory elements involved in this process, many other groups have explored how and when this mucilage is produced (Francoz et al., 2015; Voiniciuc et al., 2015c; Kreitschitz and Gorb, 2018; Yu et al., 2018).

Previous studies have identified at least 59 genes affecting seed coat cell differentiation and maturation when disrupted in *A. thaliana* (North et al., 2014; Rautengarten et al., 2014; Francoz et al., 2015; Voiniciuc et al., 2015a; Voiniciuc et al., 2015b; Griffiths et al., 2016; Ralet et al., 2016; Wardhan et al., 2016; Saez-Aguayo et al., 2017; Polko et al., 2018; Takenaka et al., 2018; Voiniciuc et al., 2018; Yu et al., 2018; Šola et al., 2019; Yang et al., 2019). These genes fall into roughly three functional 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 the TFs, and in some cases their targets, are known, there is little information about individual regulatory elements and their activity during seed coat 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 surprisingly many novel regulatory sites. Through this comparative analysis we also show that, as in animals (Thomas et al., 2011; Stergachis et al., 2013; Daugherty et al., 2017), cell lineage and developmental stage are strong determinants of the plant chromatin landscape compared to even severe environmental perturbations. This result was somewhat unexpected given that plants are so exquisitely responsive to environmental cues. Taken together, our findings call for a systematic analysis of important *A. thaliana* cell types during development and in response to major environmental cues.

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 (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 expressed at very high levels in the seed coat epidermis; it 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 to 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 the 4 and 7 DPA samples by DEseq2 (Love et al., 2014) with an adjusted p-value < 0.001 (**Figure 1A**; **Supplemental Tables 2–5**, *METHODS*). As shown in prior DNase I-seq experiments, replicates are highly reproducible, with dynamic sites clear outliers in cut count correlation plots (**Supplemental Figure 2**). We denote DHSs more accessible in 7 DPA than 4 DPA as activated DHSs, and those more accessible in 4 DPA than 7 DPA as deactivated DHSs.

Twenty-six activated DHSs resided near one of the 59 known seed coat development genes (**Supplemental Table 1**), which represents a 2.6-fold enrichment over the ten genes expected by chance (p-value < 1e6). For example, we found 7 DPA-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 transcription factor
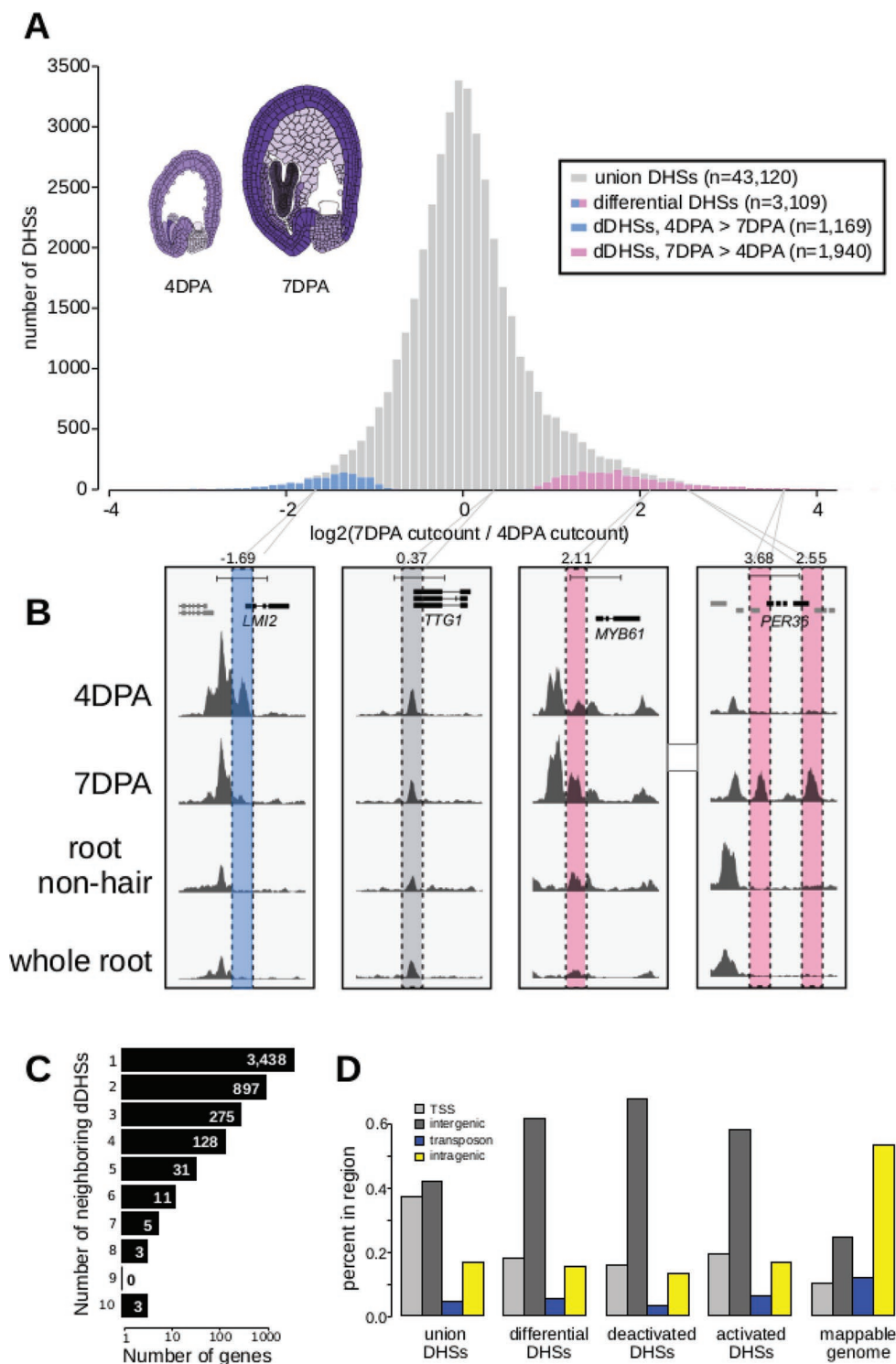
FIGURE 1 | The chromatin landscape of maturing seed coat cells. **(A)** Distribution of log2(DNase I cut count in 7 DPA/DNase I cut count in 4 DPA) for all union DNase I-hypersensitive sites (DHSs) (gray) and dynamic DHSs, with DHSs more accessible at 4 DPA appearing on the left in blue and DHSs more accessible at 7 DPA appearing on the right in pink. Diagrams of 4 DPA (left) and 7 DPA seeds (right) are shown, with purple opacity indicating *GL2* expression levels from Belmonte et al., 2013. **(B)** Examples showing a deactivated DHS, two examples of activated DHSs, and one example of a static DHS. A 5-kb region is shown in each window; all data tracks are read-depth normalized. **(C)** Distribution of the number of dynamic DHSs neighboring genes. Most genes reside next to one dynamic DHS; however, surprisingly many genes reside next to multiple dynamic DHSs. Genes neighboring dynamic DHSs are listed in **Supplemental Table 13**. **(D)** The numbers of union DHSs (uDHSs) and dynamic DHSs (dDHSs) within each genomic context: TSS, intergenic, transposon, and intragenic.
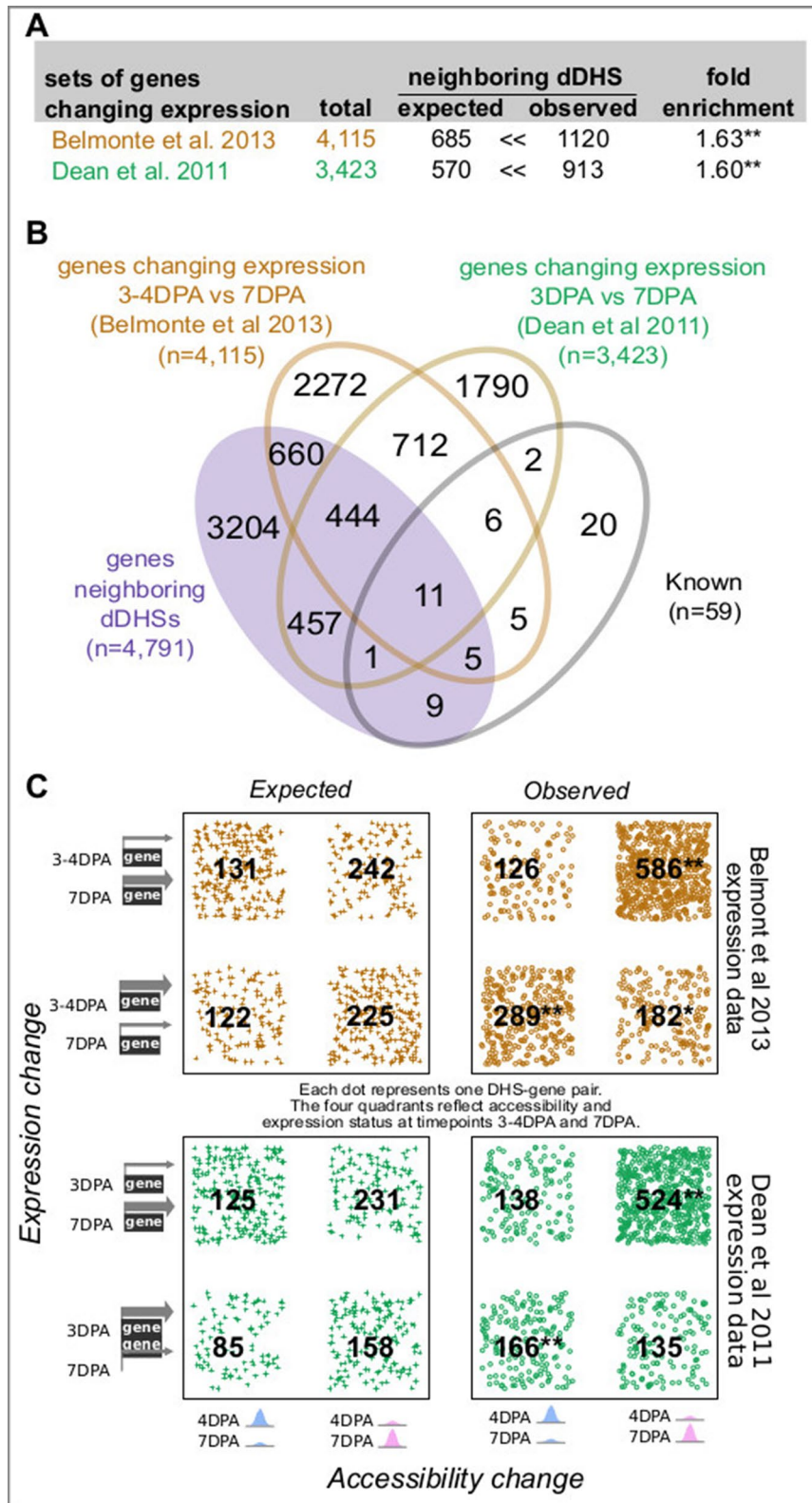
FIGURE 2 | Genes neighboring developmentally dynamic DNase I-hypersensitive sites (dDHSs) are often differentially expressed. **(A)** Overlap between the set of genes neighboring dDHSs and genes found to be differentially expressed in seed coat at stages 4 and 7 DPA in two different data sets (Dean et 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. P-values are calculated using the hypergeometric test. One asterisk (*) indicates p-value 0.01. Two asterisks (**) indicate p-value 10^{-20}.

gene, *LMI2* (Pastore et al., 2011), resides near a DHS that was deactivated during seed coat cell maturation (**Figure 1B**). Similar to previous observations (Sullivan et al., 2014), the majority of observed DHSs were static during development, such as those flanking *TTG1*, which encodes a WDR protein that regulates seed coat mucilage release (Zhang and Hülskamp, 2019) (**Figure 1B**). The regulatory landscape of seed coat cells differed significantly from the landscape of root nonhair cells, another epidermal cell type, as well as from whole roots (**Figures 1B** and **5**). Consistent with multiple regulatory inputs in development, we observed that developmentally dynamic DHSs were frequently clustered, with about a third of genes residing near more than one dynamic DHS (**Figure 1C**). For example, *LMI2*, *MYB61*, and *PER36*, shown in **Figure 1B**, all neighbor multiple DHSs but only *PER36* neighbors multiple (two) dynamic DHSs—*LMI2* and *MYB61* each neighbor only one dynamic DHS. We conclude our method detects developmentally regulated DHSs, which appear in the vicinity of known seed coat development genes and genes newly implicated in seed coat maturation.

Next, we asked whether 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 (Sullivan et al., 2014), DHSs in seed-coat-enriched cells (both dynamic and static), tended to reside in intergenic regions and near transcription start sites (TSSs, 400 bp upstream of the TSS), and were depleted in intragenic regions and transposable elements (TEs). In contrast, 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 TSSs (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 a few genes flanked by as many as ten developmentally dynamic DHS (**Figure 1C**). As we and others have shown previously, chromatin accessibility is only weakly correlated with nearby gene expression (Sullivan et al., 2015); however, dynamic chromatin accessibility (i.e. dynamic DHSs) is more frequently correlated with altered expression of nearby genes. To explore the relationship between chromatin accessibility and gene expression in maturing seeds, we took advantage of two published seed coat epidermis expression studies (Dean et al., 2011; Belmonte et al., 2013;), considering a gene to be differentially expressed if it exhibited a 2-fold expression change between developmental time points.

In the first study, Dean et al., 2011 quantified gene expression in manually dissected seed coats at 3 and 7 DPA in the Col-2

accession, identifying 3,423 genes that exhibited at least a 2-fold expression change between these developmental stages (**Figures 2A, B**; **Supplemental Figures 3A, B**). 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 accession using laser capture micro dissection. For our analysis, we used the seed coat and embryo proper expression values from globular (~3–4 DPA), heart (~4–5 DPA) and linear cotyledon (~7 DPA) stage seeds; the former approximating the 4 DPA stage while the latter approximates the 7 DPA stage (Le et al., 2010). A total of 4,115 genes exhibited at least a 2-fold expression change in seed coat. Both studies used microarrays to evaluate gene expression (**Figures 2A, B**; **Supplemental Figures 3A, B**).

For both data sets, genes with changing expression in seed coat between 4 and 7 DPA stage were significantly more likely to reside near one or more dynamic DHSs (**Figure 2**). Furthermore, increased chromatin accessibility was significantly associated with increased expression levels at both the 4 and 7 DPA stage (**Figure 2C**). Conversely, decreased chromatin accessibility was associated with decreased expression levels; however, this association was not always statistically significant (**Figure 2C**).

Although 4 DPA seeds are mainly in the globular stage of development, some will have progressed to the heart stage (Le et al., 2010; Chen et al., 2015). The INTACT transgene promoter (GL2) is activated in the embryo of both heart (4–5 DPA) and linear cotyledon (7 DPA) stage seeds. Therefore, we also examined the relationship of dynamic DHSs with genes differentially expressed between the heart and linear cotyledon stage seeds within both seed coat and embryo proper (**Supplemental Figures 3A, B**). As with the globular vs. linear cotyledon stage comparison, differentially expressed genes in seed coats were significantly more likely to reside near one or more dynamic DHS (1.58-fold). Genes differentially expressed in embryo proper were somewhat less, albeit significantly, likely (1.17-fold) to reside near one or more dynamic DHS.

We next explored whether genes neighboring multiple dynamic DHSs were enriched in gene sets previously identified to be involved in seed coat development as well as in genes with differential expression in the aforementioned studies. 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 3C**). This tendency was particularly visible for the smaller set of 59 genes with known roles in seed development, pointing to the presence of multiple DHSs as support for possible functional relevance. Thirteen of the 59 known annotated genes are differentially expressed in the Belmonte and/or Dean set but did not neighbor dynamic DHSs. However, the average number of union DHSs neighboring these 13 genes was more than twice as high as that for all genes (6.4 vs. 3.0 union DHSs, respectively, **Supplemental Figure 3D**). The magnitude of the change in expression level also modestly increased with the number of neighboring dDHSs, although this effect was only significant in the Belmonte set (**Supplemental Figure 3E**).

Genes Near Dynamic DHSs Are Implicated in Seed Coat Biology

To test whether the genes that resided near dynamic DHSs were involved in known seed coat biology, we analyzed their GO terms using GOstats (**Figure 3**; **Supplemental Tables 7** and **8**). Genes residing near deactivated DHSs were enriched for development,

regulation, response, and pigment genes. Genes nearest to activated DHSs were enriched in genes related to transport, cell wall, biosynthetic process, and localization, consistent with the known developmental processes occurring at this stage and the annotations for the 26 known seed coat development genes that resided near activated DHSs (**Supplemental Table 1**).
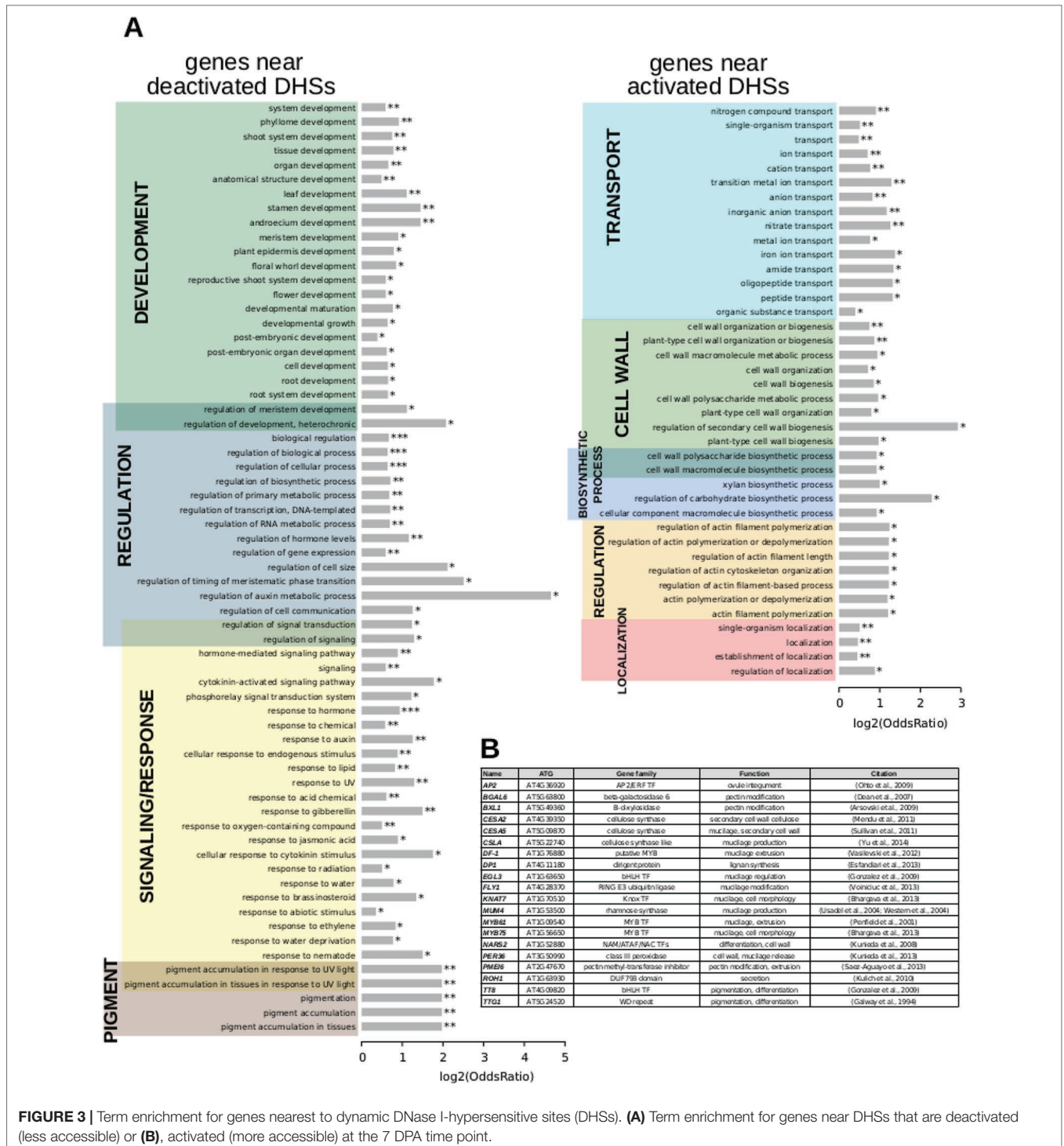
FIGURE 3 | Term enrichment for genes nearest to dynamic DNase I-hypersensitive sites (DHSs). **(A)** Term enrichment for genes near DHSs that are deactivated (less accessible) or **(B)**, activated (more accessible) at the 7 DPA time point.

Motif Families in Activated and Deactivated DHSs Are Distinct

To determine candidate transcription factors driving dynamic DHSs in seed coat development, we examined transcription factor motif enrichments, comparing developmentally dynamic DHSs to union DHSs, excluding dynamic DHSs, using AME (McLeay and Bailey, 2010). Motifs for different TF families were enriched in activated versus deactivated DHSs compared to union DHSs. Specifically, many bHLH and TCP motifs were significantly enriched in deactivated DHSs (**Figure 4A**). Motifs for many more transcription factor families were enriched in activated DHSs, including ARID, bZIP, MADS, MYB, MYB-related, and ZFHD motifs, with the majority of motifs belonging to either MYB or MYB-related transcription factors (**Figure 4B**). Previous functional studies validate our motif findings, lending support for novel associations of transcription factor motifs with seed coat development. For example, TCP3 overexpression leads to ovule integument growth defects and ovule abortion (Wei et al., 2015). In cotton, TCPs contribute to fiber elongation; cotton fibers, like seed coat cells, arise from the ovule outer integument. MYB61 is required for mucilage deposition and extrusion (Penfield et al., 2001), and AGL13, in the MADS transcription factor family, is predicted to regulate seed development (Ziegler et al., 2019).

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 from thirteen diverse plant samples primarily reflected ontogeny; in contrast, treatment with major plant hormones and/or severe stress (Sullivan et al., 2014) changed the regulatory landscape to a much lesser degree (**Figure 5A**). For example, the regulatory landscape of light-grown 7-day old seedlings inhabited a clade together with those of other light-grown seedlings that were either exposed to a severe heat shock or the plant hormone auxin. Both treatments are known to cause dramatic but drastically different changes in gene expression; yet, these did not suffice to obscure the commonalities in the regulatory landscapes of light-grown seedlings. Similarly, dark-grown seedlings (Sullivan et al., 2014), which differ profoundly in development from light-grown seedlings, clustered together. On a finer scale, the regulatory landscape of dark-grown seedlings exposed to the light-mimicking plant hormone brassinazole (BRZ) clustered closely with that of seedlings exposed to light for 24 h before harvest, whereas the landscapes of seedlings exposed to shorter light treatments before harvest and seedling grown in the dark only were more distant. Overall, the regulatory landscapes of seedling tissue, both light and dark-grown were more similar to one another than those of the two epidermal cell types included in the analysis. The regulatory landscapes enriched for seed coat cells differed profoundly from those found in root hair and nonhair cells (Sullivan et al., 2014). This tendency is also evident in a Principal Component Analysis biplot, showing the sample vectors projected on the PC1-PC2 plane (**Figure 5B**). Our results are consistent with a meta study showing that expression profiles differ more among different tissues than among tissue-controlled treatments (Aceituno et al., 2008).

In animals and humans, each sampled cell type, tissue, or condition yields novel DHSs (Stergachis et al., 2013). Published studies in plants typically only sample a limited number of conditions or tissues, falling short of denoting comprehensive regulatory landscapes. We first determined which sample pairs yielded the most dynamic DHSs (**Figure 6**). Comparing the seed-coat enriched samples to one another yielded many more dynamic DHSs than any other comparison. The regulatory landscapes for the terminally differentiated root hair and root nonhair cells yielded the lowest number of dynamic DHSs.

For analyzing all 13 samples together, we merged their DHSs, excluding those below a certain cut count (marked in gray in **Figure 6**), thereby generating 46,891 union high-confidence DHSs, covering 10,374,430 bases or ~7.4% of the genome (see *METHODS* for details). We then excluded each of the thirteen samples individually, assessing how many hypersensitive bases unique to the sample were lost. The seed coat-enriched samples (both 4 and 7 DPA) contributed the most sample-specific hypersensitive bases, followed by those found in whole roots (**Figure 7A**). Of the hypersensitive bases identified in the seed-coat-enriched samples, over half (2,858,990 bps/5,573,620 bps) were not present in 7-day-old light-grown seedlings, and over 25% (1,418,070 bps/5,573,620 bps) were not present in any of the other eleven samples examined. As more and more samples are tested, the number of identified hypersensitive base pairs is expected to plateau. We observe this phenomenon already with the 13 samples included (**Figure 7B**). Note, however, that our analysis underestimates overall DHS frequency due to subsampling all samples to the lowest read-coverage sample (14 million reads, see *METHODS*). Increasing read coverage increases the number of identified hypersensitive base pairs up to a saturation point, which depends on genome size. For the small genomes of *A. thaliana* and *Drosophila melanogaster*, this saturation point is reached with ~20 million reads for a given sample; using 14 million reads will identify ~70% of the DHSs identified with 20 million reads.

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 changed reproducibly in accessibility.

DHSs are a hallmark of regulatory DNA and thus dynamic DHSs often reside in close proximity to genes with changing expression. However, it is well-established that the association between chromatin accessibility, even if dynamic, and nearby gene expression is imperfect for several reasons (Sullivan et al., 2015). First, regulatory DNA is often poised, i.e. bound
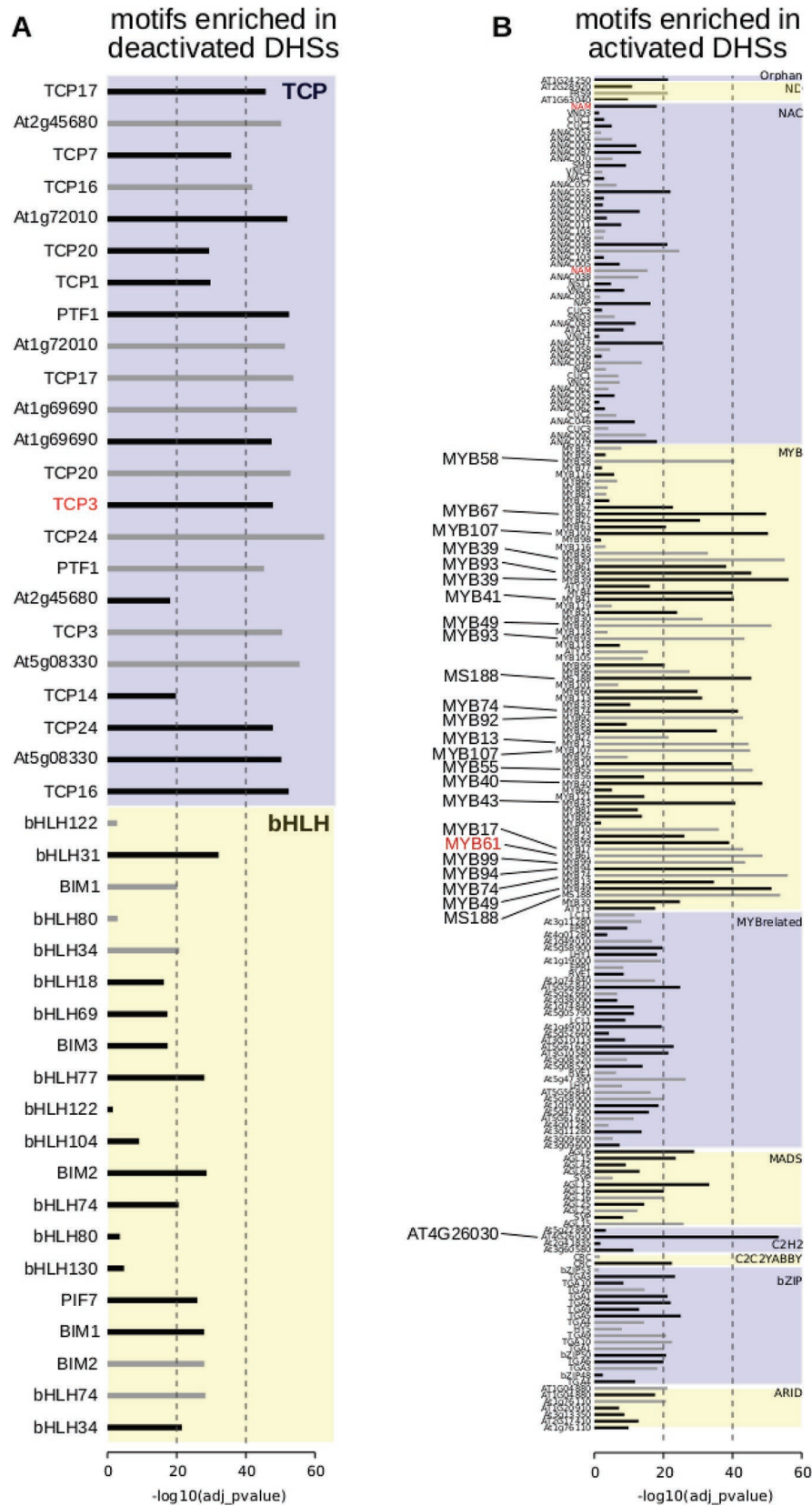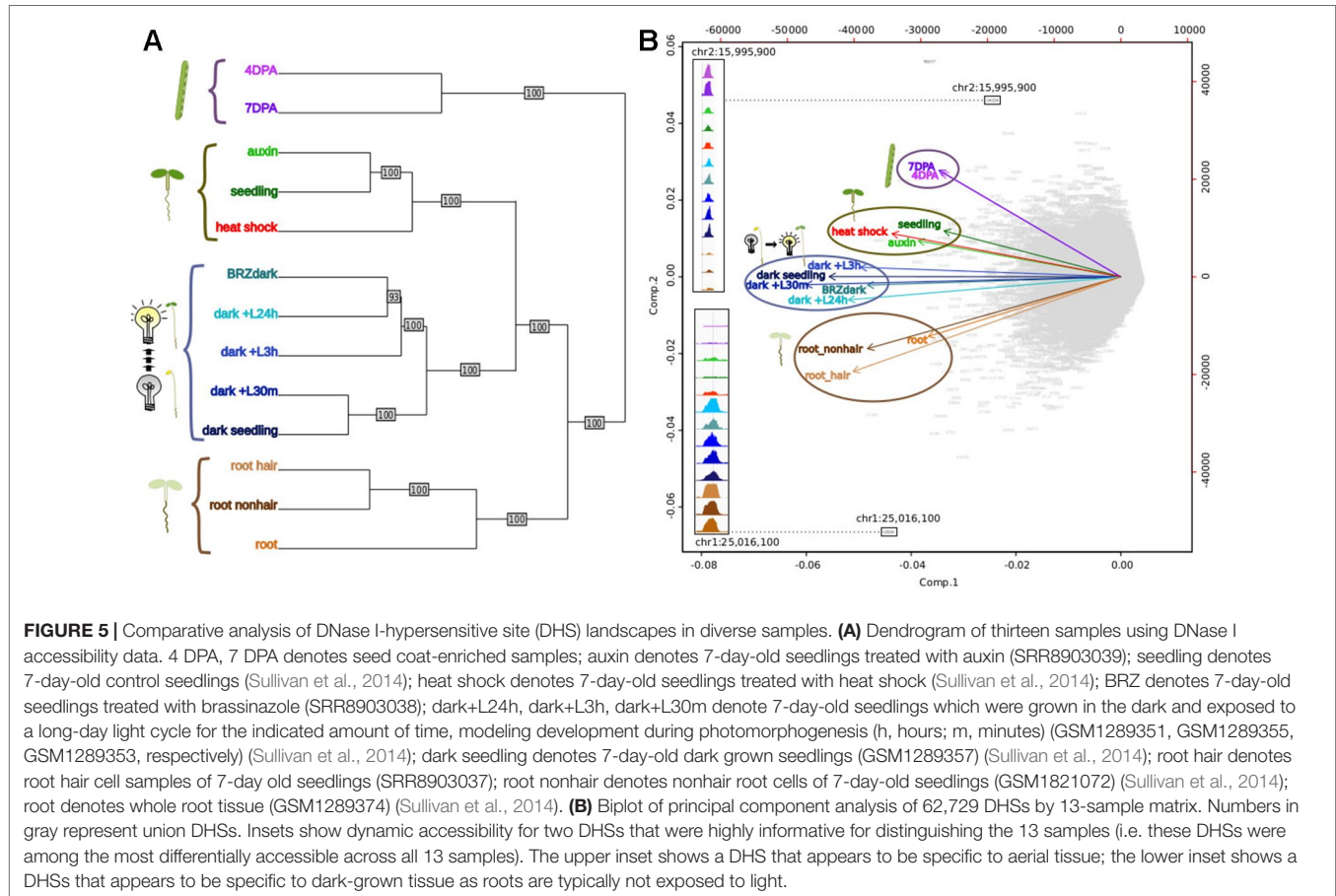
FIGURE 4 | Continued

FIGURE 5 | Comparative analysis of DNase I-hypersensitive site (DHS) landscapes in diverse samples. **(A)** Dendrogram of thirteen samples using DNase I accessibility data. 4 DPA, 7 DPA denotes seed coat-enriched samples; auxin denotes 7-day-old seedlings treated with auxin (SRR8903039); seedling denotes 7-day-old control seedlings (Sullivan et al., 2014); heat shock denotes 7-day-old seedlings treated with heat shock (Sullivan et al., 2014); BRZ denotes 7-day-old seedlings treated with brassinazole (SRR8903038); dark+L24h, dark+L3h, dark+L30m denote 7-day-old seedlings which were grown in the dark and exposed to a long-day light cycle for the indicated amount of time, modeling development during photomorphogenesis (h, hours; m, minutes) (GSM1289351, GSM1289355, GSM1289353, respectively) (Sullivan et al., 2014); dark seedling denotes 7-day-old dark grown seedlings (GSM1289357) (Sullivan et al., 2014); root hair denotes root hair cell samples of 7-day old seedlings (SRR8903037); root nonhair denotes nonhair root cells of 7-day-old seedlings (GSM1821072) (Sullivan et al., 2014); root denotes whole root tissue (GSM1289374) (Sullivan et al., 2014). **(B)** Biplot of principal component analysis of 62,729 DHSs by 13-sample matrix. Numbers in gray represent union DHSs. Insets show dynamic accessibility for two DHSs that were highly informative for distinguishing the 13 samples (i.e. these DHSs were among the most differentially accessible across all 13 samples). The upper inset shows a DHS that appears to be specific to aerial tissue; the lower inset shows a DHSs that appears to be specific to dark-grown tissue as roots are typically not exposed to light.

by transcription factors and hence accessible, without transcription occurring (Elgin, 1988); in addition, DHSs often remain accessible after transcription has occurred (Groudine and Weintraub, 1982). Second, the binding of both activators (Morgan et al., 1987) and repressors (Baniahmad et al., 1990) can remodel chromatin locally causing increased accessibility. Therefore, increases in chromatin accessibility do not necessarily translate into increases in gene expression. Finally, distal regulatory elements, i.e. enhancers residing in intergenic regions, can function at long distances and are agnostic to orientation (Banerji et al., 1981). Compared to union DHSs, we found that more dynamic, differentially accessible DHSs in seed coat-enriched cells resided in intergenic regions. As we assigned DHSs to target genes based on proximity, we will have missed long-range interactions, possibly assigning incorrect target genes. Nevertheless, we observed considerable agreement between the direction of changes in chromatin accessibility and changes in expression for neighboring genes.

Despite these limitations, dynamic DHSs are potentially useful for identifying new candidate genes that control seed coat development; moreover, their motif enrichments can point to the TFs that drive the observed DHS and gene expression dynamics. Genes near deactivated DHSs (up in 4 DPA) were associated with development, signaling, pigment, and regulation, consistent with the processes occurring during seed maturation. Genes near activated DHSs (up in 7 DPA) were associated with secretion, 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. Although most differentially expressed genes resided in close proximity to only one dynamic DHS, several hundred genes neighbored multiple dynamic DHSs, consistent with multiple regulatory inputs during development. Genes neighboring multiple dynamic DHSs were enriched for genes with altered expression in seed coat development. This trend was most strongly observed in known
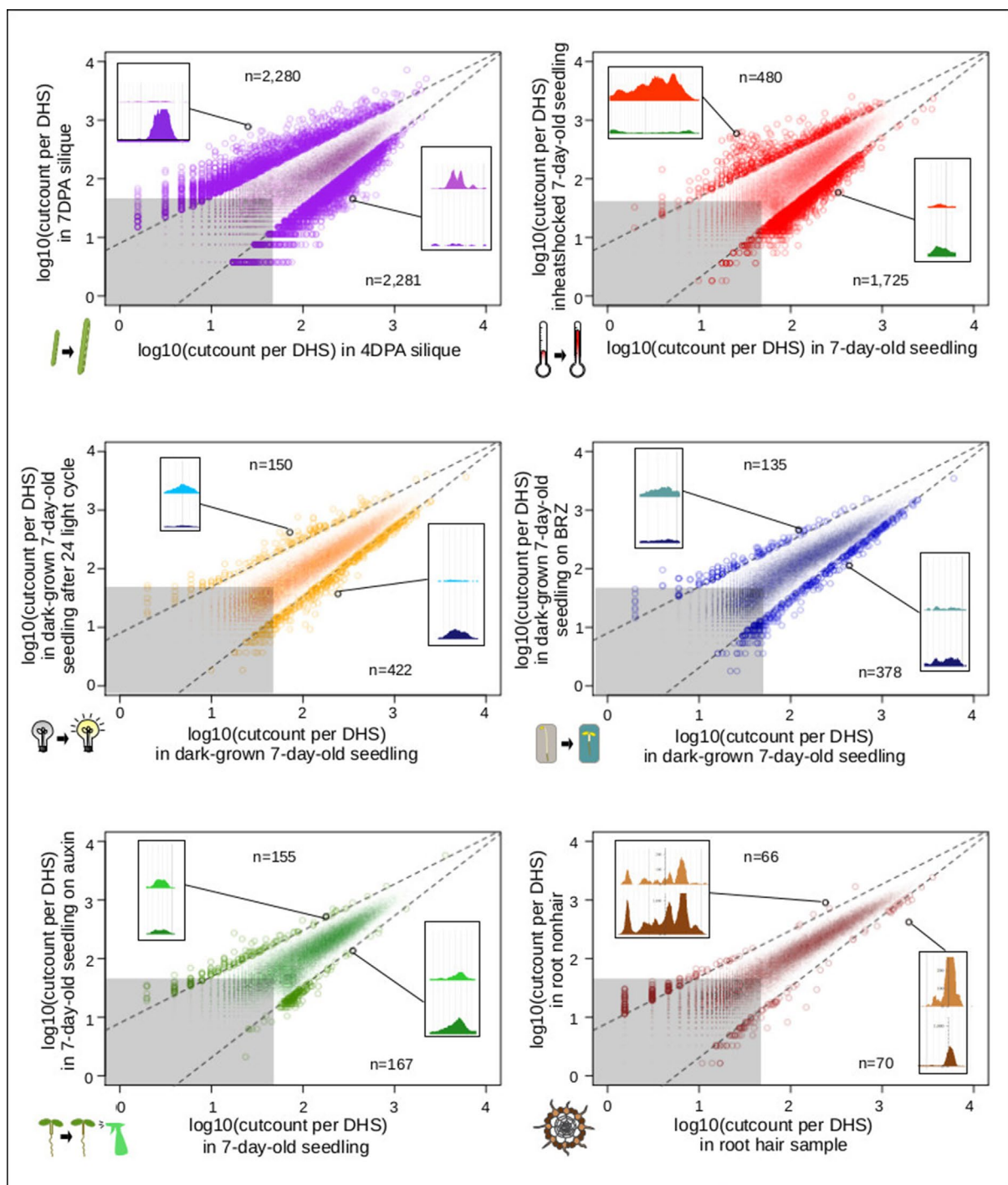
FIGURE 6 | Comparison of seed coat-enriched samples (4 and 7 DPA) results in the highest number of developmentally dynamic DNase I-hypersensitive sites (DHSs) identified among all pairs examined. Scatterplots of log10(cut counts per union DHS) for six pairwise comparisons. Dotted lines creating a cone 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 dotted lines. Screenshot insets in each graph showing an example dynamic DHSs above and below dotted lines are the following DHSs, respectively: {4 vs. 7 DPA: 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. heat shocked seedling: chr4:7,338,681–7,342,041, chr2:18,374,201–18,374,371; dark-grown seedling vs. dark-grown seedling exposed to 24-h light cycle: chr3:6,023,601–6,023,871, chr5:5,968,041–5,968,291}.
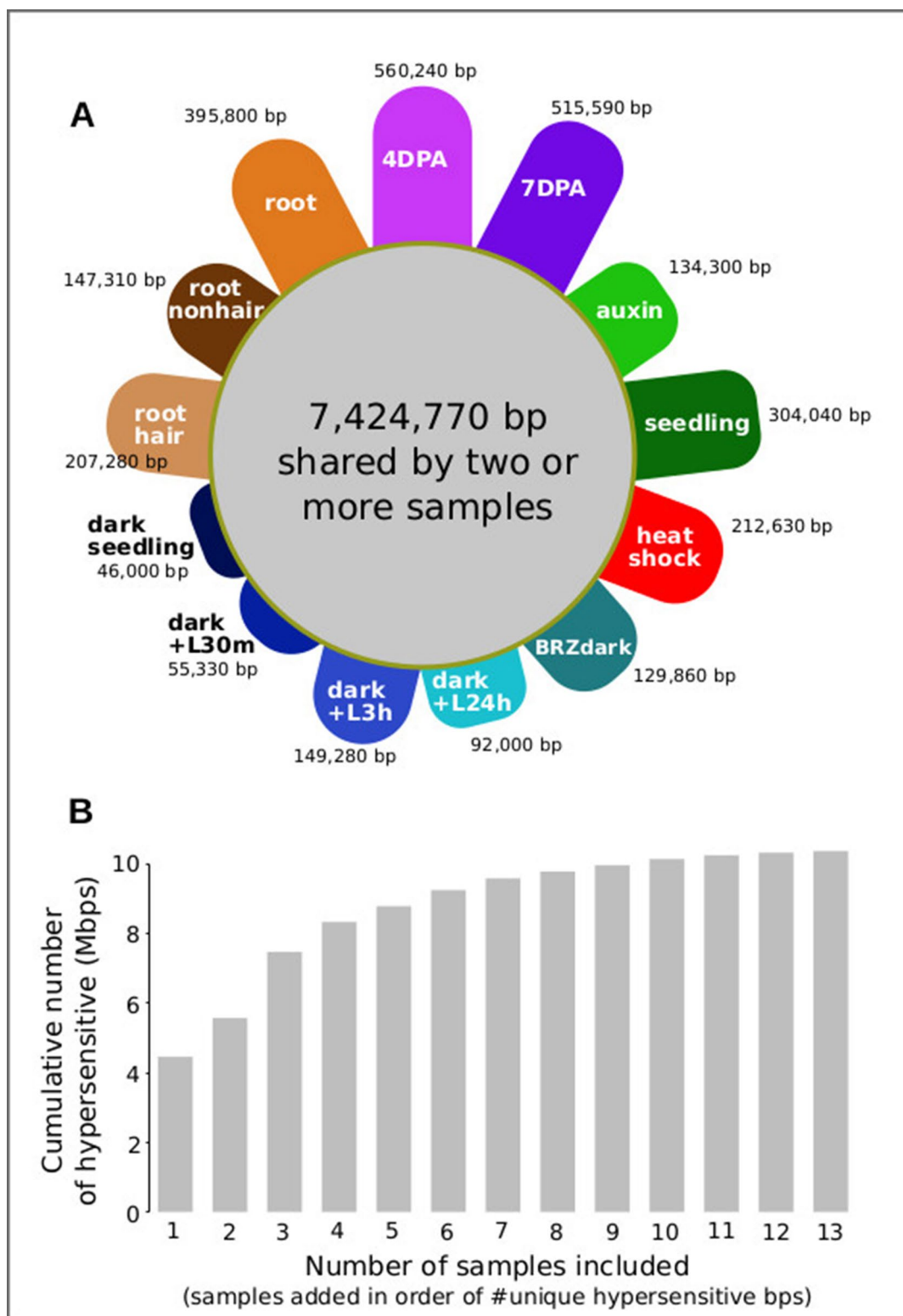
FIGURE 7 | Seed coat-enriched samples contribute the largest number of novel hypersensitive bases in a diverse set of samples. **(A)** Colored petals denote number of unique hypersensitive base pairs in each sample, gray circle denoted hypersensitive base pairs shared by two or more samples. Sample labels as in **Figure 5**; samples are grouped by seed coat-enriched samples, light-grown seedlings, dark-grown seedlings, and root samples. **(B)** Cumulative number of hypersensitive sites plateaus. Graph was generated by adding samples based on their number of unique hypersensitive base pairs, starting with the largest (4 DPA) and ending with the smallest (dark seedling).

seed coat development genes. We have noted previously that genes conditionally expressed in response to abiotic treatments tend to neighbor multiple DHSs (Alexandre et al., 2017). It appears that multiple DHSs are also a feature of developmentally dynamic genes.

Motif enrichments within activated and deactivated DHSs revealed distinct transcription factor families and individual transcription factors that may be regulating seed coat maturation. Among the TF motifs most enriched in deactivated DHSs were those of the TCP family. TCPs are involved in many aspects of development, particularly in land plants in which the class has greatly diversified (Martín-Trillo and Cubas, 2010). Consistent with its significant motif enrichment in deactivated DHSs, overexpression of *TCP3* leads to ovule integument growth defects and ovule abortion in *A. thaliana* (Wei et al., 2015). Altered expression of the most famous member of the TCP TF family, the maize TF *tb1*, contributes to the morphological changes in shoot architecture that differentiate wild teosinte and domesticated maize (Clark et al., 2006).

Among TF motifs most enriched in 7 DPA-activated DHSs were those of the MYB family. This class of TFs, present throughout Eukarya, plays important roles in plant development and stress responses (Ambawat et al., 2013). All of the MYB TFs with enriched motifs in activated DHSs belonged to the same subfamily, the R2R3 MYBs, which are involved in secondary metabolism and cell fate establishment (Stracke et al., 2001). MYB61, whose motif is enriched in our analysis, is required for mucilage production and secretion in cell coat cells (Penfield et al., 2001). Zinc finger, MADS-box, and AT-hook TFs were also enriched in 7 DPA-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).

This foray into cell-type–specific regulatory landscapes in plants, an approach that has been previously pioneered in humans and animal models and indeed has been the primary mode of analysis in these systems demonstrates the dramatic coverage and knowledge gains by analyzing specific cell types and their developmental dynamics rather than using whole seedlings or easily dissected tissues. Specifically, a single whole seedling sample previously yielded 34,288 DHSs covering ~4% of the *A. thaliana* genome (Sullivan et al., 2014). Our combined analysis of seed coat cells and 11 other samples generated a set of 46,891 union DHSs which accounted for ~7.4% of the *A. thaliana* genome. Of these, 1,978 were entirely non-overlapping with DHSs in the other 11 samples. Expressed in base pairs this result appears even more impressive: of 10,374,430 hypersensitive, accessible bps in all 13 samples, 560,240 hypersensitive bps (> 5%) were unique to the seed coat-enriched samples. This result demonstrates that cell-type–specific DHS profiling holds enormous promise for expanding our knowledge of the *A. thaliana* regulatory landscape. Although heat stress, auxin, and BRZ treatments cause dramatic changes in genes expression, our comparative analysis shows that cell lineage and developmental stage rather these treatments are reflected in regulatory landscapes, which is consistent with prior knowledge of poised transcription factors (Elgin, 1988), in particular those occupying heat shock promoters (Vihervaara

et al., 2018). Our findings argue for exploring regulatory landscapes across all plant cell types, across development, and in response to relevant conditions to fully understand how chromatin accessibility and gene expression are integrated into precise expression patterns. 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).

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 (Western et al., 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 two 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 (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 20× objective. Image of seed coat cell layer (**Supplemental Figure 1A**) is a single image using the 63× water immersion objective.

Data Processing for Seed Coat Analysis

Five DNase I-seq libraries, including biological replicates for each time point, were sequenced and aligned to the TAIR10 reference genome using bwa/0.6.2. Because number of peaks called is a function of read depth, 24 million reads mapping to chromosomes 1 to 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 (4 DPA-DS20201 and 7 DPA-DS21306). These 24M-read bam files were used to call DHSs (peaks) using the HOTSPOT program (John et al., 2011). DHSs from these two samples were merged to create a union set of 43,120 DHSs. DESeq2 (Love et al., 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 to 5, excluding centromeres, from all five samples (4 DPA-DS20201, 4 DPA-DS20131, 4 DPA-DS20132, 7 DPA-DS21306, 7 DPA-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 3 and 7 DPA; these genes were used for integration with dynamic DHS data. The genes that change expression by two or more folds in Belmonte et al., 2013 were extracted from the published normalized expression data (Dataset S2 in the Belmonte et al., 2013 publication). We used the hypergeometric test to measure how different the observed number of DHS-gene pairs in certain configurations were compared to the expected number. For example, there were 2,131 genes that had 2-fold more expression at 7 DPA than 3 to 4 DPA in the Belmonte et al., 2013 data set, and 3,269 genes that were near dDHSs that were more accessible at 7 DPA than 4 DPA. Given that there are 28,775 genes total, we expect 2,131 × 3,269/28,775 ≈ 242 DHS-gene pairs with this configuration if accessibility and expression are randomly associated. We observe 586 such DHS-gene pairs, which is a statistically significant excess (p-value < 10^{-20}).

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 < 0.001 are shown in **Figure 3**.

Motif Enrichment

Enrichment of motifs (O'Malley et al., 2016) in sequence underlying the 1,182 deactivated DHSs (dDHSs more accessible at 4 DPA than 7 DPA) and the 2,258 activated DHSs (dDHSs less accessible at 4 DPA than 7 DPA) as compared to the sequences underlying the 39,680 union DHSs, excluding dynamic DHSs, was evaluated using AME version 5.0.5 with the rank-sum test (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 < 0.01 are listed in **Supplemental Tables 9** and **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 to 5, excluding centromeres (chr1:13,698,788–15,897,560; chr2: 2,450,003–5,500,000; chr3:11,298,763–14,289,014; chr4:1,800,002–5,15 0,000; 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., 2011), 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 by creating 100 trees from random subsamples of 10,000 DHSs using the ape package (Paradis et al., 2004). Principal Component Analysis was performed on the 62,729 by 13 matrix. For the PCA, we excluded nine DHSs within the first 50 kb 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 7**).

Pairs of Samples Resulting in Dynamic 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 (Sullivan et al., 2014).

DATA AVAILABILITY STATEMENT

AUTHOR CONTRIBUTIONS

AA designed the experiments, and together with AT executed DNase I-seq experiments. AS and KB performed data analysis and made the figures. KB, RS, RT, SN, AJ, and SS assisted with the bioinformatics analysis and data processing. PS, FN, MW, and MD sequenced DNase I-treated samples. AS, KB and CQ wrote the manuscript. JS and JN facilitated experiments and assisted in writing the manuscript. All authors read, commented on, and approved the manuscript.

FUNDING

ACKNOWLEDGMENTS

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2019.01434/full#supplementary-material

SUPPLEMENTAL FIGURE 1 | Confocal microscopy of INTACT-tagged nuclei in seed coat epidermis. **(A)** Confocal of whole seed at 4DPA from the INTACT line GL2pro:NTF/ACT2pro: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 GL2pro:NTF/ACT2pro:BirA (Deal and Henikoff, 2010). GFP-fluorescing nuclei are readily observable in the outer most layer of the seed coat. Scale is 100 um.

SUPPLEMENTAL FIGURE 2 | Correlation of normalized cut counts within 150bp windows. The three replicates of one time point (4DPA) had pairwise Pearson's correlation coefficients (PCCs) of 0.96 and 0.97. The two replicates of the other time point (7DPA) had a 0.92 Pearson's correlation coefficient. PCCs between time points were lower (maximum PCC=0.89), with multiple obvious outliers corresponding to regions of differential accessibility (light purple: regions identified in subsequent analysis as more accessible in 4DPA; dark purple: regions identified in subsequent analysis as more accessible in 7DPA). Plots display the normalized cut counts within 100,000 150 bp sliding windows (130 bp overlap) in the first 2,000,045 bp of chromosome 1.

SUPPLEMENTAL FIGURE 3 | 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)** The set of genes neighboring multiple dynamic DHSs tend to contain more genes related to seed coat development than expected at random. This is seen in the set of 59 known seed coat development genes (**Supplemental Table 1**) as well as in genes with differential expression (Dean et al. 2011; Belmonte et al. 2013). Dashed line indicates Fold Enrichment of 1, which indicates no enrichment over random expectation. Significance of difference in Fold Enrichment compared to random expectation indicated by asterisks: *p-value<0.05, ***p-value<1e6. **(D)** Distribution of number of union DHSs neighboring all genes. The number of union DHSs neighboring the 13 known genes that are differentially expressed in the Belmonte and/or Dean set but did not neighbor dynamic DHSs are indicated by the placement of colored dots along the x-axis. Of these 13 genes, five were identified only in the Belmonte set (AT1G79840, AT3G13540, AT5G67360, AT2G37260, and AT1G21070), two only in the Dean set (AT5G67030 and AT3G09820), and six in both sets (AT5G23940, AT1G02720, AT4G36890, AT3G15510, AT2G18280, and AT5G35550). **(E)** The magnitude of the change in expression of genes neighboring 1, 2, and 3-or-more dynamic DHSs. Orange boxplots show expression change in the Belmonte et al. 2013 data set. Green boxplots show expression changes in the Dean et al. 2011 data set. The difference in mean between the abs(log2(fold change)) for genes neighboring 2 and 3-or-more dynamic DHSs and those neighboring 1 dynamic DHSs is small but significant for the Belmonte et al. 2013 data set, but not for the Dean et al. 2011 data set (p-values: 2.5e-4, 5e-7, 0.111,0.125).

REFERENCES

Aceituno, F. F., Moseyko, N., Rhee, S. Y., and Gutiérrez, R. A. (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. doi: 10.1186/1471-2164-9-438

Alexandre, C. M., Urton, J. R., Jean-Baptiste, K., Huddleston, J. L., Dorrity, M. W., Cuperus, J. T., et al. (2017). Complex relationships between chromatin accessibility, sequence divergence, and gene expression in A. Thaliana. *Mol. Biol. Evol.* 35 (4), 837–854. doi: 10.1093/molbev/msx326.

Ambawat, S., Sharma, P., Yadav, N. R., and Yadav, R. C. (2013). MYB transcription factor genes as regulators for plant responses: an overview. *Physiol. Mol. Biol. Plants: Int. J. Funct. Plant Biol.* 19 (3), 307–321. doi: 10.1007/s12298-013-0179-1

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

Banerji, J., Olson, L., and Schaffner, W. (1983). A lymphocyte-specific cellular enhancer is located downstream of the joining region in immunoglobulin heavy chain genes. *Cell* 33 (3), 729–740. doi: 10.1016/0092-8674(83)90015-6

Baniahmad, A., Steiner, C., Köhne, A. C., and Renkawitz, R. (1990). Modular structure of a chicken lysozyme silencer: involvement of an unusual thyroid

hormone receptor binding site. *Cell* 61 (3), 505–514. doi: 10.1016/0092-8674(90)90532-J

Belmonte, M. F., Kirkbride, R. C., Stone, S. L., Pelletier, J. M., Bui, A. Q., Yeung, E. C., et al. (2013). Comprehensive developmental profiles of gene activity in regions and subregions of the arabidopsis seed. *Proc. Natl. Acad. Sci. U.S.A.* 110 (5), E435–E444. doi: 10.1073/pnas.1222061110

Carlson, M. (2016) ath1121501.db: Affymetrix Arabidopsis ATH1 Genome Array annotation data (chip ath1121501). R package version 3.2.3.

Chen, L.-Q., Lin, I. W., Qu, X.-Q., Sosso, D., Heather, E., Londoño, M. A., et al. (2015). A cascade of sequentially expressed sucrose transporters in the seed coat and endosperm provides nutrition for the arabidopsis embryo. *Plant Cell* 27 (3), 607–619. doi: 10.1105/tpc.114.134585

Chung, J. H., Bell, A. C., and Felsenfeld, G. (1997). Characterization of the chicken β-globin insulator. *Proc. Natl. Acad. Sci.* 94 (2), 575–580. doi: 10.1073/pnas.94.2.575

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

Copenhaver, G. P., and Pikaard, C. S. (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. *Plant J. Cell Mol. Biol.* 9 (2), 259–272. doi: 10.1046/j.1365-313X.1996.09020259.x

Daugherty, A. C., Yeo, R., Buenrostro, J. D., and Greenleaf, W. J. (2017). Chromatin accessibility dynamics reveal novel functional enhancers in C. elegans. *Gen. Res.* 27 (12), 2096–2107. doi: 10.1101/088732

Deal, R. B., and Henikoff, S. (2010). A simple method for gene expression and chromatin profiling of individual cell types within a tissue. *Dev. Cell* 18 (6), 1030–1040. doi: 10.1016/j.devcel.2010.05.013

Dean, G., Cao, Y., Xiang, D., Nicholas, J., Ramsay, P. L., Ahad, A., et al. (2011). Analysis of gene expression patterns during seed coat development in arabidopsis. *Mol. Plant* 4 (6), 1074–1091. doi: 10.1093/mp/ssr040

Elgin, S. C. (1988). The formation and function of DNase I hypersensitive sites in the process of gene activation. *J. Biol. Chem.* 263 (36), 19259–19262.

Esfandiari, E., Jin, Z., Abdeen, A., Griffiths, J. S., Western, T. L., and Haughn, G. W. (2013). Identification and analysis of an outer-seed-coat-specific promoter from Arabidopsis Thaliana. *Plant Mol. Biol.* 81 (1-2), 93–104. doi: 10.1007/s11103-012-9984-0

Falcon, S., and Gentleman, R. (2007). Using GOstats to test gene lists for go term association. *Bioinformatics* 23 (2), 257–258. doi: 10.1093/bioinformatics/btl567

Francoz, E., Ranocha, P., Burlat, V., and Dunand, C. (2015). Arabidopsis seed mucilage secretory cells: regulation and dynamics. *Trends Plant Sci.* 20 (8), 515–524. doi: 10.1016/j.tplants.2015.04.008

García-Fayos, P., Bochet, E., and Cerdà, A. (2010). Seed removal susceptibility through soil erosion shapes vegetation composition. *Plant Soil* 334 (1-2), 289–297. doi: 10.1007/s11104-010-0382-6

Garwood, N. C. (1985). The role of mucilage in the germination of cuipo, cavanillesia platanifolia (H. & B.) H. B. K. (Bombacaceae), a tropical tree. *Am. J. Bot.* 72 (7), 1095–1105. doi: 10.1002/j.1537-2197.1985.tb08357.x

Griffiths, J. S., Crepeau, M.-J., Ralet, M.-C., Seifert, G. J., and North, H. M. (2016). Dissecting seed mucilage adherence mediated by FEI2 and SOS5. *Front. Plant Sci.* 7 (June), 1073. doi: 10.3389/fpls.2016.01073

Groudine, M., and Weintraub, H. (1982). Propagation of globin DNAase I-hypersensitive sites in absence of factors required for induction: a possible mechanism for determination. *Cell* 30 (1), 131–139.

Gutterman, Y., and Shem-Tov, S. (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 J. Plant Sci.* 45 (4), 317–327. doi: 10.1080/07929978.1997.10676695

Honma, T., and Goto, K. (2001). Complexes of MADS-box proteins are sufficient to convert leaves into floral organs. *Nature* 409 (6819), 525–529. doi: 10.1038/35054083

John, S., Sabo, P. J., Canfield, T. K., Lee, K., Vong, S., Weaver, M., et al. (2013). "Genome-scale mapping of DNase I hypersensitivity," in *Current Protocols in Molecular Biology / Edited by Frederick M. Ausubel [et Al.]* Chapter 27 (July): Unit 21.27. doi: 10.1002/0471142727.mb2127s103

John, S., Sabo, P. J., Thurman, R. E., Sung, M.-H., Biddie, S. C., Johnson, T. A., et al. (2011). Chromatin accessibility pre-determines glucocorticoid receptor binding patterns. *Nat. Genet.* 43 (3), 264–268. doi: 10.1038/ng.759

Kreitschitz, A., and Gorb, S. N. (2018). The micro- and nanoscale spatial architecture of the seed mucilage-comparative study of selected plant species. *PloS One* 13 (7), e0200522. doi: 10.1371/journal.pone.0200522

Kunieda, T., Mitsuda, N., Ohme-Takagi, M., Takeda, S., Aida, M., Tasaka, M., et al. (2008). NAC family proteins NARS1/NAC2 and NARS2/NAM in the outer integument regulate embryogenesis in arabidopsis. *Plant Cell* 20 (10), 2631–2642. doi: 10.1105/tpc.108.060160

Le, B. H., Cheng, C., Bui, A. Q., Javier, A., Wagmaister, K. F., Pelletier, H. J., et al. (2010). Global analysis of gene activity during arabidopsis seed development and identification of seed-specific transcription factors. *Proc. Natl. Acad. Sci. U.S.A.* 107 (18), 8063–8070. doi: 10.1073/pnas.1003530107

Lin, X., Kaul, S., Rounsley, S., Shea, T. P., Benito, M. I., Town, C. D., et al. (1999). Sequence and analysis of chromosome 2 of the plant Arabidopsis Thaliana. *Nature* 402 (6763), 761–768. doi: 10.1038/45471

Love, M. I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-Seq Data with DESeq2. *Genome Biol.* 15 (12), 550. doi: 10.1186/s13059-014-0550-8

Martín-Trillo, M., and Cubas, P. (2010). TCP Genes: a family snapshot ten years later. *Trends Plant Sci.* 15 (1), 31–39. doi: 10.1016/j.tplants.2009.11.003

McLeay, R. C., and Bailey, T. L. (2010). Motif enrichment analysis: a unified framework and an evaluation on ChIP data. *BMC Bioinformatics* 11 (1), 165. doi: 10.1186/1471-2105-11-165

Morgan, W. D., Williams, G. T., Morimoto, R. I., Greene, J., Kingston, R. E., and Tjian, R. (1987). Two transcriptional activators, CCAAT-box-binding transcription factor and heat shock transcription factor, interact with a human hsp70 gene promoter. *Mol. Cell. Biol.* 7 (3), 1129–1138.

Neph, S., Kuehn, M. S., Reynolds, A. P., Haugen, E., Thurman, R. E., Johnson, A. K., et al. (2012). BEDOPS: high-performance genomic feature operations. *Bioinformatics* 28 (14), 1919–1920. doi: 10.1093/bioinformatics/bts277

North, H. M., Berger, A., Saez-Aguayo, S., and Ralet, M.-C. (2014). Understanding polysaccharide production and properties using seed coat mutants: future perspectives for the exploitation of natural variants. *Ann. Bot.* 114 (6), 1251–1263. doi: 10.1093/aob/mcu011

O'Malley, R. C., Huang, S.-S. C., Song, L., Lewsey, M. G., Bartlett, A., Nery, J. R., et al. (2016). Cistrome and epicistrome features shape the regulatory DNA landscape. *Cell* 165 (5), 1280–1292. doi: 10.1016/j.cell.2016.04.038

Paradis, E., Claude, J., and Strimmer, K. (2004). APE: analyses of phylogenetics and evolution in R Language. *Bioinformatics* 20 (2), 289–290. doi: 10.1093/bioinformatics/btg412

Pastore, J. J., Limpuangthip, A., Yamaguchi, N., Wu, M.-F., Sang, Y., Han, S.-K., et al. (2011). LATE MERISTEM IDENTITY2 acts together with leafy to activate apetala1. *Development* 138 (15), 3189–3198. doi: 10.1242/dev.063073

Penfield, S., Meissner, R. C., Shoue, D. A., Carpita, N. C., and Bevan, M. W. (2001). MYB61 is required for mucilage deposition and extrusion in the arabidopsis seed coat. *Plant Cell* 13 (12), 2777–2791. doi: 10.1105/tpc.010265

Pinyopich, A., Ditta, G. S., Savidge, B., Liljegren, S. J., Baumann, E., Wisman, E., et al. (2003). Assessing the redundancy of MADS-box genes during carpel and ovule development. *Nature* 424 (6944), 85–88. doi: 10.1038/nature01741

Polko, J. K., Barnes, W. J., Voiniciuc, C., Doctor, S., Steinwand, B., Hill Jr, J. L., et al. (2018). SHOU4 proteins regulate trafficking of cellulose synthase complexes to the plasma membrane. *Curr. Biol.* 28 (19), 3174–82.e6. doi: 10.1016/j.cub.2018.07.076

Ralet, M.-C., Crépeau, M.-J., Vigouroux, J., Tran, J., Berger, A., Sallé, C., et al. (2016). Xylans provide the structural driving force for mucilage adhesion to the Arabidopsis seed coat. *Plant Physiol.* 171 (1), 165–178. doi: 10.1104/pp.16.00211

Rautengarten, Ca, Ebert, B., Moreno, I., Temple, H., Herter, T., Link, B., et al. (2014). The Golgi localized bifunctional UDP-rhamnose/UDP-galactose transporter family of Arabidopsis. *Proc. Natl. Acad. Sci. U.S.A.* 111 (31), 11563–11568. doi: 10.1073/pnas.1406073111

Saez-Aguayo, S., Rautengarten, C., Temple, H., Sanhueza, D., Ejsmentewicz, T., Sandoval-Ibañez, O., et al. (2017). UUAT1 Is a Golgi-Localized UDP-uronic acid transporter that modulates the polysaccharide composition of arabidopsis seed mucilage. *Plant Cell* 29 (1), 129–143. doi: 10.1105/tpc.16.00465

Šola, K., Gilchrist, E. J., Ropartz, D., Wang, L., Feussner, I., Mansfield, S. D., et al. (2019). RUBY, a putative galactose oxidase, influences pectin properties and promotes cell-to-cell adhesion in the seed coat epidermis of arabidopsis. *Plant Cell* 31 (4), 809–831. doi: 10.1105/tpc.18.00954

Stergachis, A. B., Neph, S., Reynolds, A., Humbert, R., Miller, B., Paige, S. L., et al. (2013). Developmental fate and cellular maturity encoded in human regulatory DNA landscapes. *Cell* 154 (4), 888–903. doi: 10.1016/j.cell.2013.07.020

Stracke, R., Werber, M., and Weisshaar, B. (2001). The R2R3-MYB gene family in arabidopsis thaliana. *Curr. Opin. Plant Biol.* 4 (5), 447–456. doi: 10.1016/S1369-5266(00)00199-0

Sullivan, A. M., Arsovski, A. A., Lempe, J., Bubb, K. L., Weirauch, M. T., Sabo, P. J., et al. (2014). Mapping and dynamics of regulatory dna and transcription factor networks in a. thaliana. *Cell Rep.* 8 (6), 2015–2030. doi: 10.1016/j.celrep.2014.08.019

Sullivan, A. M., Bubb, K. L., Sandstrom, R., Stamatoyannopoulos, J. A., and Queitsch, C. (2015). DNase I Hypersensitivity mapping, genomic footprinting, and transcription factor networks in plants. *Curr. Plant Biol.* 3-4 (September), 40–47. doi: 10.1016/j.cpb.2015.10.001

Takenaka, Y., Kato, K., Ogawa-Ohnishi, M., Tsuruhama, K., Kajiura, H., Yagyu, K., et al. (2018). Pectin RG-I rhamnosyltransferases represent a novel plant-specific glycosyltransferase family. *Nat. Plants* 4 (9), 669–676. doi: 10.1038/s41477-018-0217-7

Talbot, D., Collis, P., Antoniou, M., Vidal, M., Grosveld, F., and Greaves, D. R. (1989). A dominant control region from the human beta-globin locus conferring integration site-independent gene expression. *Nature* 338 (6213), 352–355. doi: 10.1038/338352a0

Thomas, S., Li, X.-Y., Sabo, P. J., Sandstrom, R., Thurman, R. E., Canfield, T. K., et al. (2011). Dynamic reprogramming of chromatin accessibility during drosophila embryo development. *Genome Biol.* 12 (5), R43. doi: 10.1186/gb-2011-12-5-r43

Thurman, R. E., Rynes, E., Humbert, R., Vierstra, J., Maurano, M. T., Haugen, E., et al. (2012). The accessible chromatin landscape of the human genome. *Nature* 489 (7414), 75–82. doi: 10.1038/nature11232

Tominaga-Wada, R., Iwata, M., Sugiyama, J., Kotake, T., Ishida, T., Yokoyama, R., et al. (2009). The GLABRA2 homeodomain protein directly regulates CESA5 and XTH17 gene expression in arabidopsis roots. *Plant J. Cell Mol. Biol.* 60 (3), 564–574. doi: 10.1111/j.1365-313X.2009.03976.x

Vihervaara, A., Duarte, F. M., and Lis, J. T. (2018). Molecular mechanisms driving transcriptional stress responses. *Nat. Rev. Genet.* 19 (6), 385–397. doi: 10.1038/s41576-018-0001-6

Voiniciuc, C., Engle, K. A., Günl, M., Dieluweit, S., Schmidt, M. H.-W., Yang, J.-Y., et al. (2018). Identification of key enzymes for pectin synthesis in seed mucilage. *Plant Physiol.* 178 (3), 1045–1064. doi: 10.1104/pp.18.00584

Voiniciuc, C., Günl, M., Schmidt, M. H.-W., and Usadel, B. (2015a). Highly Branched Xylan Made by IRREGULAR XYLEM14 and MUCILAGE-RELATED21 Links Mucilage to Arabidopsis Seeds. *Plant Physiol.* 169 (4), 2481–2495. doi: 10.1104/pp.15.01441

Voiniciuc, C., Schmidt, M. H.-W., Berger, A., Yang, B., Ebert, B., Scheller, H. V., et al. (2015b). MUCILAGE-RELATED10 produces galactoglucomannan that maintains pectin and cellulose architecture in arabidopsis seed mucilage. *Plant Physiol.* 169 (1), 403–420. doi: 10.1104/pp.15.00851

Voiniciuc, C., Yang, B., Schmidt, M. H.-W., Günl, M., and Usadel, B. (2015c). Starting to gel: how arabidopsis seed coat epidermal cells produce specialized secondary cell walls. *Int. J. Mol. Sci.* 16 (2), 3452–3473. doi: 10.3390/ijms16023452

Wardhan, V., Pandey, A., Chakraborty, S., and Chakraborty, N. (2016). Chickpea Transcription Factor CaTLP1 interacts with protein kinases, modulates ROS accumulation and promotes aba-mediated stomatal closure. *Sci. Rep.* 6 (December), 38121. doi: 10.1038/srep38121

Wei, B., Zhang, J., Pang, C., Yu, H., Guo, D., Jiang, H., et al. (2015). The molecular mechanism of sporocyteless/nozzle in controlling Arabidopsis ovule development. *Cell Res.* 25 (1), 121–134.

Western, T. L., Skinner, D. J., and Haughn, G. W. (2000). Differentiation of mucilage secretory cells of the arabidopsis seed coat. *Plant Physiol.* 122 (2), 345–356. doi: 10.1104/pp.122.2.345

Windsor, J. B., Symonds, V. V., Mendenhall, J., and Lloyd, A. M. (2000). Arabidopsis seed coat development: morphological differentiation of the outer integument. *Plant J. Cell Mol. Biol.* 22 (6), 483–493. doi: 10.1046/j.1365-313x.2000.00756.x

Wu, C., Bingham, P. M., Livak, K. J., Holmgren, R., and Elgin, S. C. (1979a). The chromatin structure of specific genes: I. evidence for higher order domains of defined DNA sequence. *Cell* 16 (4), 797–806. doi: 10.1016/0092-8674(79)90095-3

Wu, C., Wong, Y. C., and Elgin, S. C. (1979). The chromatin structure of specific genes: II. disruption of chromatin structure during gene activity. *Cell* 16 (4), 807–814. doi: 10.1016/0092-8674(79)90096-5

Yang, B., Voiniciuc, C., Fu, L., Dieluweit, S., Klose, H., and Usadel, B. (2019). TRM4 is essential for cellulose deposition in arabidopsis seed mucilage by maintaining cortical microtubule organization and interacting with CESA3. *New Phytol.* 221 (2), 881–895. doi: 10.1111/nph.15442

Yang, X., Dong, M., and Huang, Z. (2010). Role of mucilage in the germination of artemisia sphaerocephala (Asteraceae) achenes exposed to osmotic stress and salinity. *Plant Physiol. Biochem.: PPB / Societe Francaise Physiol. Vegetale* 48 (2-3), 131–135. doi: 10.1016/j.plaphy.2009.12.006

Yang, X., Zhang, W., Dong, M., Boubriak, I., and Huang, Z. (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), e24346. doi: 10.1371/journal.pone.0024346

Yu, L., Lyczakowski, J. J., Pereira, C. S., Kotake, T., Yu, X., Li, A., et al. (2018). The patterned structure of galactoglucomannan suggests it may bind to cellulose in seed mucilage. *Plant Physiol.* 178 (3), 1011–1026. doi: 10.1104/pp.18.00709

Zabidi, M. A., Arnold, C. D., Schernhuber, K., Pagani, M., Rath, M., Frank, O., et al. (2015). Enhancer-core-promoter specificity separates developmental and housekeeping gene regulation. *Nature* 518 (7540), 556–559. doi: 10.1038/nature13994

Zhang, B., and Hülskamp, M. (2019). Evolutionary analysis of mbw function by phenotypic rescue in arabidopsis thaliana. *Front. Plant Sci.* 10 (March), 375. doi: 10.3389/fpls.2019.00375

Ziegler, D. J., Khan, D., Kalichuk, J. L., Becker, M. G., and Belmonte, M. F. (2019). Transcriptome Landscape of the Early Brassica Napus Seed. *J. Integr. Plant Biol.* 61 (5), 639–650. doi: 10.1111/jipb.12812