ATAC-seq Reveals Chromatin Accessibility Changes During Maize Seed Development

Minghua Li ✉

Biotechnology Research Center, Cuixi Academy of Biotechnology, Zhuji, 311800, China

✉ Corresponding author: minghua.li@cuixi.org

Preferred citation for this article:

Li M.H., 2025, ATAC-seq reveals chromatin accessibility changes during maize seed development, Maize Genomics and Genetics, 16(4): 182-201 (doi: 10.5376/mgg.2025.16.0017)

Abstract Chromatin accessibility plays a key role in regulating gene expression during plant development. While transcriptional changes during maize seed development have been well studied, how chromatin accessibility shifts over time is still not fully understood. This study used ATAC-seq (Assay for Transposase-Accessible Chromatin using sequencing) to examine open chromatin regions (OCRs) across the maize genome at several seed developmental stages. By combining these data with RNA-seq results, we tracked how accessibility patterns change as seeds grow and found that these changes are closely tied to gene activity. Most OCRs were located near promoters or enhancers. Motif analysis pointed to several transcription factor families—such as bZIP, MYB, and NAC—as likely players in developmental control. Functional annotation showed that OCR-associated genes are highly enriched in pathways like starch biosynthesis, hormone signaling, and embryo development. This study maps how chromatin opens and closes across maize seed development, identifies potential regulatory elements and key TFs, and offers useful insights and targets for epigenetic studies and molecular breeding in maize.

Keywords Maize; ATAC-seq; Chromatin accessibility; Seed development; Regulatory network

1 Introduction

In the last few decades, scientists have found that maize seed development is controlled by many systems. These systems work together in different parts of the plant and at different times. At the start of seed development, the embryo begins to form. Several special types of endosperm cells also start to appear. These include aleurone cells, cells that store starch, and transfer cells. The way these cells form depends on when and where certain genes are turned on. This process is mainly controlled by transcription factors (TFs). These are proteins that help turn genes on or off.

Let's look at the endosperm as an example. Two TFs called NAKED ENDOSPERM1 and 2 (NKD1/2), which belong to the AL2/GL2-type group, are needed to make sure aleurone cells form the right way. These proteins help turn on genes related to cell identity, how the cell responds to hormones, and how it stores nutrients. In maize, NKD1/2 affect over 6% of the endosperm's gene activity. One important gene they influence is *Opaque2*, which helps make storage proteins. Another TF, Opaque11, acts like a main switch. It controls other TFs such as DOF3 and Opaque2 (Gontarek et al., 2016; Zhan et al., 2018). It also helps regulate genes that produce starch and proteins. In the embryo, a set of TFs called LAFL proteins help the seed move from early growth to the mature stage. For example, a B3-type TF known as ABSCISIC ACID-INSENSITIVE 3, along with similar proteins in maize, turns on genes needed in the late stages of embryo development. It also activates genes that protect the seed from drying out. All these TFs don't work alone. They interact closely with plant hormones. Auxin, which comes from the central cell, helps start the endosperm's development. Later, when ABA (abscisic acid) levels go up, it helps the seed mature and prepares it for dormancy (Yang et al., 2022). Gibberellins are also part of this process. If something in this system fails, the seed might grow abnormally or end up with the wrong size or content.

Apart from DNA sequences and how available TFs are, the structure of chromatin—how tightly or loosely DNA is packed—also plays a big part in whether a gene is active or silent. Chromatin that's "open," meaning DNA is

more exposed and not wrapped tightly by nucleosomes, often overlaps with key DNA control sites known as cis-regulatory elements (CREs). These include promoters and enhancers that serve as landing pads for TFs and other chromatin-related proteins. In plants, large-scale mapping has shown that many of these open areas are located near gene promoters, and the more open a promoter is, the more likely that gene is being used. A good example is ATAC-seq, a method that detects open DNA. Studies using ATAC-seq in plants reveal that about 15%-30% of open regions sit close to transcription start sites. Genes near these regions tend to be turned on, while those far away often need enhancer regions—parts of the genome that can loop over long distances to reach and affect a gene (Lu et al., 2016; Zhang et al., 2022).

Chromatin accessibility is an important epigenetic feature that regulates gene expression during plant development. Although transcriptional changes during maize seed development have been extensively studied, the dynamic changes in chromatin accessibility have not been deeply analyzed. In this study, we systematically analyzed the genome-wide open chromatin regions (OCRs) in multiple developmental stages of maize seeds using ATAC-seq (assay for transposase-accessible chromatin sequencing) technology, and combined RNA-seq transcriptome data to reveal the dynamic changes in chromatin accessibility with developmental stages, which were highly correlated with gene expression activity. We found that OCRs were mainly enriched in promoter and enhancer regions, and motif analysis further identified multiple transcription factor families such as bZIP, MYB, and NAC that may be involved in developmental regulation. Functional annotation indicated that OCR-related genes were significantly enriched in key biological processes such as starch synthesis, hormone signaling, and embryo development. This study has drawn a panoramic map of chromatin accessibility during maize seed development, revealing potential regulatory elements and key regulatory factors, providing a theoretical basis and target resources for maize epigenetic research and molecular breeding.

2 Chromatin Accessibility Changes during Seed Developmental Stages

2.1 Temporal changes in OCR number and abundance

Maize seeds possess thousands of genomic regions that become accessible (or inaccessible) as development proceeds. We identified on the order of 40 000-60 000 high-confidence OCRs at each developmental stage, indicating that a substantial portion of the maize genome is involved in regulatory activity during seed formation. In the very early stages immediately post-fertilization, when the endosperm is still a coenocyte and the embryo is just beginning to divide, we detected the fewest OCRs (roughly 3.8×10^4). As development advanced into the cellularization of the endosperm and differentiation of embryonic organs, the number of OCRs increased markedly – peaking at nearly 6.0×10^4 open regions by the middle of the seed-filling phase. This trend suggests an overall opening of the chromatin landscape as the seed's cellular complexity and metabolic activity ramp up. The expansion in OCR count may reflect the activation of many tissue-specific genes and cis-elements required for storage product biosynthesis, cell expansion, and other processes that dominate mid-development. After this peak, we observed that the total number of OCRs plateaued or even modestly declined in later stages of maturation (e.g. approaching desiccation), hinting that chromatin accessibility might contract again as seed development concludes. A similar pattern of initial increase and later reduction in accessible sites has been reported in other plant developmental contexts, such as maize leaf maturation and inflorescence development.

2.2 Genomic distribution of accessibility changes (promoters, enhancers, intergenic regions)

We next examined where in the genome these OCRs are located and how their genomic distribution changes with development. Consistent with prior studies in plants, we found that promoter regions (defined here as a few hundred base pairs upstream of transcription start sites) are highly enriched for open chromatin. Across stages, roughly 20%–25% of all OCRs were located in promoters or genic upstream regions. These promoter-proximal OCRs include sites like core promoters and 5′ UTR regulatory sequences that are accessible when the associated gene is active. We observed that many genes which are broadly expressed during seed development (e.g. genes for core metabolism or cell maintenance) have constitutively accessible promoters at all stages, whereas genes with stage-specific expression often showed promoter accessibility only at the corresponding stage. For instance, the promoter of a gene encoding a starch synthase enzyme was open (ATAC-seq peak present) during the mid-to-late endosperm development when starch was actively being synthesized, but not accessible at the earliest embryonic

stage. This suggests a tight coupling of promoter chromatin state with developmental gene activation. Besides promoters, a large fraction of OCRs (~50% or more) resided in intergenic or intronic regions, often far from gene transcription start sites. Many of these distal OCRs likely represent enhancer elements or other distal CREs that regulate gene expression through long-range interactions. We found numerous examples where an intergenic OCR could be assigned (by proximity or through known QTL) to a particular gene and developmental process. A notable case is the distal open region upstream of the UNBRANCHED3 (*UB3*) gene in maize inflorescences: this region, known as KRN4, is an enhancer that influences *UB3* expression and hence kernel row number (Du et al., 2020). In our seed dataset, the orthologous region corresponding to KRN4 showed accessible chromatin specifically in inflorescence tissues but not in developing seeds, consistent with its specialized role in ear development (and validating our method's tissue-specific resolution). Conversely, we identified distal OCRs within the seed that could function as enhancers controlling seed-specific genes. Many such OCRs were found near genes involved in hormone biosynthesis/signaling, storage accumulation, or developmental transitions, suggesting the presence of enhancers coordinating these gene networks.

Interestingly, some intergenic OCRs were located within transposable element (TE) sequences or flanking them. It has been noted that TEs contribute to regulatory innovation in maize by donating sequences that can act as enhancers. We observed stage-specific accessibility at certain MITE and LTR retrotransposon sequences in the seed genome, hinting that they might serve as developmental enhancers or chromatin boundary elements (Bubb et al., 2024). Indeed, transposon-derived enhancers have played roles in maize evolution (e.g., Hopscotch at tb1) and likely also in development. Globally, while the absolute number of promoter OCRs changed somewhat over time (tracking the number of expressed genes), the proportion of OCRs in promoters vs. distal regions remained relatively stable across stages. However, the specific identity of accessible intergenic regions changed markedly: early in development one set of enhancers is open (e.g. those for embryo patterning genes), whereas later, a different set of enhancers (e.g. for storage product genes) becomes accessible. These observations underscore that dynamic enhancer usage is a feature of seed development. Our findings align with prior single-cell ATAC-seq results in maize, which showed that roughly one-third of detected OCRs are cell-type or stage specific – reflecting enhancers that turn on or off depending on developmental context. Overall, the developmental stage of the seed can be "read out" in terms of its chromatin accessibility profile: certain promoter and enhancer regions are hallmarks of specific stages, reinforcing the idea that chromatin accessibility changes are an integral part of developmental gene regulation.

2.3 Genome-wide visualization of chromatin accessibility dynamics
To understand the changes in chromatin accessibility during the growth of maize seeds, we examined the open chromatin regions (OCRs) across the entire genome. We tracked which regions were open or closed at different stages. One of our methods was to display the ATAC-seq signal trajectories of different chromosomal regions at different time points. This made it easier for us to identify regions where chromatin accessibility increased or decreased over time. For example, we used a genome browser to examine a 2 Mb region on chromosome 1. In this region, some peaks remained strong throughout all stages. These stable peaks were mostly near the constitutively active housekeeping genes. However, we also found some peaks that only appeared at certain time points. In a region containing a group of storage protein genes, the signal was almost non-existent three days after pollination (DAP3). But by 8 to 12 days after pollination, we saw clear and strong peaks. This was in good agreement with the time when these genes began to be active in the endosperm. Therefore, by observing the ATAC-seq signals over time, we can see when certain regions of the genome become more open, which usually matches the activation time of nearby genes (Zheng et al., 2025).

We also created heatmaps to display the ATAC-seq read levels around all open chromatin regions (OCRs). The samples were arranged by developmental stage to facilitate the observation of accessibility changes over time. Some regions were open early but closed later. Others remained closed initially and only opened later. Some regions remained open throughout. These different patterns were clearly shown in the heatmaps. Next, we classified the OCRs into different types. We called them "early open", "late open", and "always open". Early open OCRs were usually close to genes related to cell division and tissue growth. These genes were typically active

when the seeds were still young. In contrast, late open OCRs were mostly close to genes related to nutrient storage and desiccation. These are common functions in mature seeds. Therefore, the timing and location of chromatin opening seem to match the activities of nearby genes at different stages.

These time-based patterns weren't just visible in the heatmap. We also ran a principal component analysis (PCA) on the ATAC-seq profiles, which clearly separated the samples along the first two axes. Developmental stages were spread out in order from earliest to latest along the first component, showing that most of the variation in chromatin accessibility comes from the seed's progression through time. From this, we could outline a sort of "pseudotime" map based on accessibility patterns that mirrors the actual developmental process of the seed.

3 Association between Open Chromatin and Gene Expression
3.1 Integration of OCR and transcriptome data

By aligning ATAC-seq peaks with gene annotations and comparing to RNA-seq expression levels, we found a strong global correspondence between chromatin accessibility at gene promoters and the activation of those genes. In general, genes that were highly expressed in a given developmental stage had an accessible promoter in that stage, whereas genes that were transcriptionally silent tended to lack ATAC-seq peaks at their promoters. A positive correlation between promoter openness and gene expression was evident when plotting accessibility scores against mRNA levels for all genes: most points fell along an upward trend, indicating that increased chromatin accessibility is associated with higher transcription. For example, at 10 DAP (a stage of active endosperm filling), genes with the strongest ATAC-seq signals at their promoters (top decile of accessibility) showed on average a several-fold higher RNA-seq read count than genes with no detectable promoter accessibility. This broad correlation was also observed within gene clusters and pathways. Take the starch biosynthesis pathway: multiple enzymes in this pathway (such as ADP-glucose pyrophosphorylase, starch synthases, starch branching enzymes) are co-expressed during endosperm filling, and accordingly we detected open chromatin at the promoters of many of their genes specifically at the stages when starch is accumulating. One notable example is the *Sh2* gene (encoding ADP-glucose pyrophosphorylase large subunit) – its promoter ATAC-seq signal is weak at early stages and becomes very strong at mid/late development, paralleling a jump in *Sh2* transcript levels (consistent with *Sh2*'s known role in starch production in the endosperm) (Figure 1) (Yu et al., 2023). Such correspondence suggests that the establishment of open chromatin at these promoters is a key regulatory event enabling the transcription of storage product genes. To quantify the relationship, we computed Spearman's correlation between promoter accessibility (ATAC-seq read count in the promoter region) and gene expression (RNA-seq read count) across our developmental series. The majority of genes exhibited positive correlations, and for a large subset the correlation was statistically significant ($p < 0.01$). These findings align with earlier reports in maize seedlings and other plants that accessible promoters generally mark actively transcribed genes. It is worth noting that while promoter accessibility and expression are broadly coupled, our analysis also revealed some exceptions: a small subset of genes had accessible promoters but low expression, or vice versa. Overall, our integrated analysis supports the intuitive model that during seed development, when a gene's product is needed, its promoter becomes accessible (likely through chromatin remodeling events) and the gene is transcribed, whereas genes not needed remain in a closed chromatin state. This relationship provides confidence that ATAC-seq is capturing functionally relevant regulatory changes, and it allowed us to define sets of co-regulated genes based on combined chromatin–transcriptomic behavior.

We didn't just focus on promoters. We also looked at enhancers and the genes they might control during seed development. To find possible links, we used two main steps. First, we matched each open chromatin region (OCR) with the closest gene. Then, we compared how open the region was with how much the gene was expressed. Often, we saw that when an OCR became more open, the nearby gene became more active too. Many of these enhancers were found 5 to 100 kilobases away from the genes they might affect. These regions usually opened at certain times and matched changes in gene activity during development.

For example, we found an OCR about 30 kb before a gene that helps make auxin. This region opened only during the early stages of seed growth. At the same time, the auxin gene showed high expression. This suggests the OCR

might help start auxin production, which is important early on. We also saw OCRs near LEA protein genes open later in development. This matched the increase in LEA gene expression during those stages. While this doesn't prove the enhancer-gene link directly, the pattern fits what we know about plant development. It also agrees with recent 3D genome studies in maize. Those studies showed that chromatin loops can bring faraway regulatory regions close to the genes they control, especially when the genes are active (Zhou et al., 2024a).
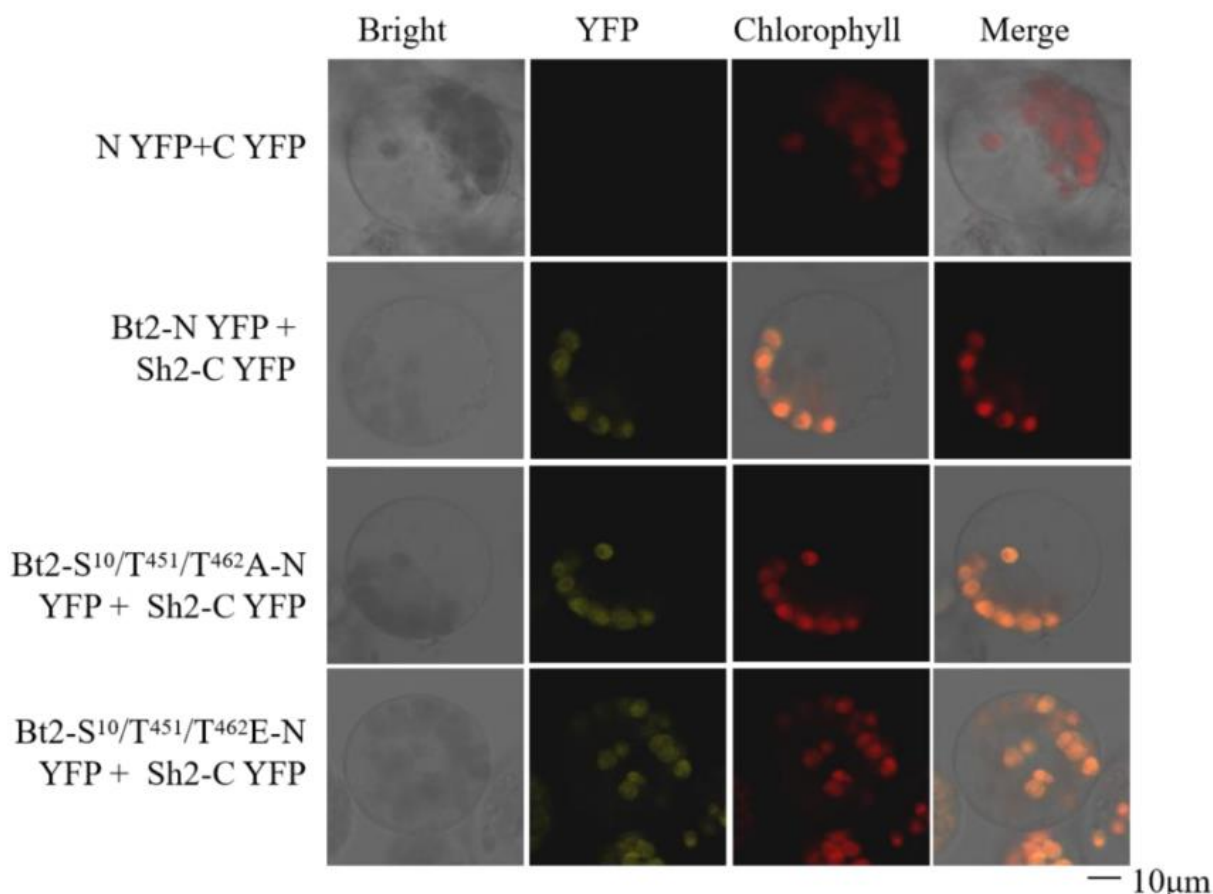

Figure 1 The bimolecular fluorescence complementation between different mutants of Bt2 and Sh2. Fluorescence signals indicate that the Bt2 protein retains its interaction with the large subunit Sh2 (Adopted from Yu et al., 2023)

3.2 Co-regulated gene clusters and functional enrichment

Once we saw that open chromatin often matches gene activity, we wanted to look closer. We tried to find sets of genes that changed in both openness and expression at the same time. We also wanted to learn what these genes do. So, we grouped them by looking at two things: how their promoter regions opened or closed over time, and how their expression levels went up or down during those same stages. This gave us a few clear gene groups that seemed to work together. For instance, one group showed open promoters and high expression early, between 0 and 3 DAP, but then slowed down later. Another group stayed quiet in the beginning but became active and open from about 8 to 15 DAP. Some genes stayed on the whole time, and others only turned on briefly. After making these clusters, we used GO term and KEGG pathway tools to check what each group might be doing. These tools helped us guess the key role of each gene set. Some were tied to early growth, while others helped with storage or stress later on. In short, grouping by openness and activity gave us a good idea of what different genes are doing at each stage.

Genes in the "early-stage" cluster (high accessibility/expression in the young seed) were significantly enriched for GO categories related to DNA replication, cell cycle, chromatin organization, and embryonic pattern formation. This makes sense because shortly after fertilization, the endosperm undergoes rapid nuclear divisions and the embryo starts forming fundamental structures. Indeed, early-stage accessible genes included those encoding core

cell-cycle regulators (e.g. cyclins, DNA polymerase subunits) and factors for morphogenesis (such as embryo sac development genes). Many of these early genes have to be tightly controlled once their window closes; accordingly, their promoters lost accessibility as development progressed and cell division slowed. In contrast, genes in the "late-stage" cluster (accessible/expressed primarily during seed filling and maturation) were enriched for processes like starch and sucrose metabolism, storage protein accumulation, abiotic stress response (desiccation tolerance), and hormone signaling. Key metabolic genes such as *Sh2* and *Bt2* (ADP-glucose pyrophosphorylase subunits for starch biosynthesis) and Zein family genes (storage proteins) fell into this category, consistent with the known timing of reserve deposition in endosperm. Their promoters became accessible in mid-development, likely under the control of endosperm master regulators (like O2 and PBF for zeins), and remained open through maturation to drive high-level expression. The late cluster was also rich in genes for LEA proteins and small heat-shock proteins (involved in dehydration tolerance), matching the seed's preparatory steps for dormancy. GO terms for response to ABA and seed maturation were significantly overrepresented here, reflecting the ABA-mediated gene activation in late seed development.

Another interesting cluster was one with biphasic or persistent expression (genes on at multiple stages). This set included many "housekeeping" genes (basic metabolism, translation, etc.) as well as some regulatory genes that act across stages. Their promoters were constitutively accessible, which is in line with their continuous requirement. GO analysis for this cluster showed enrichment in primary metabolic processes and protein synthesis, indicating these genes support general cellular functions throughout development. Notably, some of these genes still showed subtle changes in expression (e.g. upregulated during the intense growth phase), and correspondingly slight increases in ATAC signal, but overall they remained accessible at all times – a possible signature of genes that maintain the basal physiological state of the seed.

We also connected our co-regulation findings to known gene regulatory networks. For example, we observed that many genes controlled by the Opaque11 network in the endosperm (as identified by Feng et al. (2018)) fell into a common cluster: they all start being expressed around the onset of endosperm cellularization and peak during storage accumulation. In our data, these genes (including several zein storage protein genes, pyruvate phosphate dikinase for carbon metabolism, etc.) indeed had promoters that became accessible around 4–6 DAP and remained open thereafter, consistent with O11's timing of action. The enrichment analysis of that cluster highlighted terms like nutrient reservoir activity and transcription factor activity, illustrating that it contained both metabolic genes and regulatory genes co-activated in the mid-stage endosperm. Similarly, an early cluster contained both NKD1/2 and their downstream target genes (like Betl genes for basal endosperm transfer layer formation), all showing early accessibility and expression, which corresponds to NKD's role in early endosperm patterning.

3.3 Transcriptional features of active vs. repressive OCRs

While most open chromatin regions in our data were associated with gene activation, we did identify a subset of OCRs that did not correspond to actively expressed genes. These "accessible-but-silent" regions suggest that not all accessible chromatin is permissive for transcription-some may be bound by repressors or correspond to poised elements. For example, we found a number of gene promoters that were accessible (as indicated by ATAC-seq peaks) even when the gene's mRNA was either low or undetected. Closer inspection revealed that many of these genes are ones that become active only under specific conditions or later in development (outside the sampled stages), or they are tissue-specific within the seed (e.g. restricted to a cell layer with low abundance in whole-seed RNA). However, some cases were puzzling – genes with accessible promoters and apparently no expression in our dataset. Interestingly, motif analysis on these accessible-but-inactive promoters showed enrichment of binding sites for known transcriptional repressors. One notable motif was the RY cis-element (CATGCA), which is recognized by B3-domain factors of the VAL (VP1/ABI3-Like) family that recruit Polycomb complexes to silence gene expression during certain stages. Indeed, studies in Arabidopsis have shown that VAL1/VAL2 bind to DNA and recruit PRC2 (Polycomb Repressive Complex 2) to establish H3K27me3 marks, keeping seed maturation genes off until the appropriate time. We observed that some late embryogenesis genes in maize (analogous to those VAL would target) had accessible chromatin at their promoters in early development but were not yet expressed – plausibly because repressive complexes were in place. As the seed transitions to maturation, these

genes then get expressed (with activation by ABI3/VP1 factors), implying a hand-off from a repressed but open state to an active state. This scenario matches the concept of poised chromatin: the chromatin is open (perhaps to allow rapid activation potential) but transcription is held in check by repressor proteins. The VAL1/VAL2 example has been documented in Arabidopsis and our findings suggest a similar mechanism could be at play in maize seeds (Yuan et al., 2020).

Another example of gene silencing comes from distant OCRs that might work like silencers. We found several open regions in the genome, located between genes, that could be silencer elements. These spots were near genes that are usually turned off, such as imprinted genes or known negative regulators. Some of these regions were open in one tissue, like the maternal layer or aleurone, but closed in another, like the embryo or starchy endosperm. This matched the activity of nearby genes. In the tissue where the OCR was open, the linked gene was turned off. One idea is that these open regions might attract tissue-specific repressors. These could be MYB or homeodomain proteins that help shut down gene expression in one part of the seed. Here's one example. A gene that codes for a transcription factor needed for germination was active in the embryo but silent in the endosperm. We found an open DNA region upstream of this gene, but only in the endosperm. That region contained repeat sequences, which might be bound by a repressor. This suggests the endosperm might open that area just to bring in a silencing complex. That way, it keeps the germination gene turned off until the right time.

It is also worth mentioning that some genes displayed discordant chromatin–expression patterns due to possibly post-transcriptional regulation. For instance, a gene might have an open promoter and even produce transcripts, but those transcripts are not accumulating (perhaps due to RNA instability or translation being blocked). In such cases, our RNA-seq might show low steady-state mRNA despite an accessible promoter. However, these scenarios are harder to identify without further data (like nascent transcription assays or proteomics) and constituted a minority in our observations. Overall, the majority of OCRs in maize seeds act as conventional positive regulatory elements (promoters or enhancers facilitating gene expression), but a subset seem to mark regions of active chromatin that nonetheless correspond to repressed genes. The existence of these cases highlights the complexity of epigenetic regulation: an open chromatin state is generally permissive but must be interpreted in the context of bound factors. Accessible chromatin can serve as a platform for either activation or repression, depending on the proteins recruited to that site. Our data provide candidates for such regulatory phenomena in maize seed development, such as VAL-like repressor targets and potential silencer elements. Recognizing these will be important for a complete understanding of the chromatin code governing seed development – it is not a simple on/off switch, but a combinatorial platform where open DNA may need further cues to drive or block transcription. Importantly, these findings caution that strategies to modulate gene expression by targeting open chromatin (for crop improvement purposes) must account for the possible presence of repressive complexes at those sites.

4 Key Transcription Factors and Regulatory Network Construction

4.1 Cis-element discovery and transcription factor footprinting

We performed de novo motif enrichment analysis on the sequences of our OCRs to uncover over-represented DNA sequence motifs, which likely correspond to binding sites of relevant TF families. This analysis yielded a rich collection of motifs, many of which could be matched to known plant cis-elements. Strikingly, the highest enrichment scores were observed for motifs belonging to a few major TF families, consistent with those families' known roles in seed development. MIKC MADS-box motifs were among the top hits: a specific 8-bp CC[A/T]6GG sequence (the CArG-box motif) appeared frequently in accessible regions that open during early endosperm and embryo differentiation. This motif is recognized by MADS-domain transcription factors, such as AGAMOUS-like proteins. In fact, our motif resembled the binding site of Arabidopsis SEPALLATA (SEP) and SHORT VEGETATIVE PHASE (SVP) MADS factors, which are known to play roles not only in floral development but also in embryogenesis and endosperm development. The enrichment of the CArG motif in cluster III OCRs (those open at 4–6 DAP) suggests that MADS-box TFs are active in early seed development. Indeed, maize homologs of AGL15/18 (such as ZmMADS47) have been implicated in embryonic tissue development, and our data support that they bind accessible sites during those stages (Galli et al., 2018).

Another highly enriched motif was the RY-repeat (CATGCA[TG]), which is the hallmark binding site for B3 domain transcription factors of the ABI3/VAL family. We found RY motifs particularly enriched in OCRs associated with late embryogenesis genes and seed maturation genes. This aligns with the presence of B3 TFs like Viviparous1 (ZmABI3) and FUSCA3-like factors in seeds. ABI3/VP1 is known to bind RY elements to activate late maturation genes (like LEA protein genes) and also to maintain repression of germination genes during seed development. The presence of accessible RY-containing elements in our late-stage ATAC data suggests that ABI3 and related B3 TFs are engaging those sites. Interestingly, VAL repressors also recognize RY motifs, so the context (cofactor presence, chromatin marks) likely determines whether a given RY element acts as an activation site or a repression site. Our data showed RY motifs in both promoter OCRs of late genes (likely activation by ABI3) and in some early accessible but silent sites (likely binding of VAL early, then replaced by ABI3 late) – supporting the dynamic use of this motif.

Auxin-responsive elements (AuxREs) were another prominent category: the TGTCTC motif (and slight variants thereof) was enriched in OCRs during early embryo development. This sequence is the binding site for Auxin Response Factors (ARFs), which mediate auxin signaling. Auxin is a key signal at the start of seed development, and ARFs like ARF5/MONOPTEROS are crucial for embryonic patterning. The detection of AuxRE motifs in early-stage open regions aligns with the idea that auxin signaling pathways are transcriptionally active then. We also found that many of these AuxRE-containing OCRs were near genes involved in embryo axis formation and organ initiation. Supporting evidence comes from known ARF targets and regulators of auxin distribution (e.g. YUCCA auxin biosynthesis genes) which had accessible AuxRE motifs upstream when those genes were being expressed. Furthermore, using a motif clustering approach, we identified several ARF family members with potential activity: ARF25, ARF34, and ARF35 motifs (predicted by slight differences in base preferences) appeared in our embryo-specific OCR cluster. This is consistent with recent reports that specific ARFs have predominant roles in maize embryo morphogenesis.

4.2 Identification of stage-specific transcription factors

The pattern of motif enrichment in accessible chromatin provides strong clues about which transcription factors are most active at each developmental stage. To distill this information, we cross-referenced the enriched motifs and footprints with the expression patterns of the corresponding TF genes. We found a remarkable concordance: transcription factor families whose motifs were enriched in early-stage OCRs tend to have members that are highly expressed in early seed development, and similarly for mid or late stages. This enabled us to propose a set of key TFs acting at specific times:

Early seed development (0–3 DAP): Our data point to the involvement of MADS-box and auxin-related TFs. In particular, we identified a maize MADS-domain factor most similar to Arabidopsis AGL15/AGL18 that is highly expressed in the developing endosperm during cellularization, and whose binding sites (CArG motifs) are accessible at that time. This suggests it may play a role in endosperm differentiation. Another early regulator is auxin-responsive TFs: multiple ARF genes (e.g. *ZmARF34*, *ZmARF35*) show peak expression in the developing embryo around the transition from proembryo to organogenesis, and correspondingly, ARF motifs are prominent in embryo OCRs. Additionally, we saw evidence of the B3 TF LEC2 (LEAFY COTYLEDON2) or its maize ortholog being active early: the LEC2 motif (RY element) appears even in some early OCRs, and ZmLEC2 transcripts were detected in the 4–6 DAP embryo. LEC2 is known to initiate somatic embryogenesis and influence storage gene expression later, so its early presence might prime the embryo for maturation.

Mid-stage seed development (4-8 days): At this stage, the embryo begins to form its basic organs. Meanwhile, the endosperm stops dividing and starts to store nutrients. We found that some bZIP transcription factors become active during this period. One key factor is Opaque2 (O2). It begins to appear in the maize endosperm after 6 to 8 days. O2 helps to activate genes encoding storage proteins. We also found that its binding site (TGACGT, a part of the P-box) is more common in open chromatin regions (OCRs) that become accessible after 6 to 10 days. Another important factor we discovered is ABI19. Like VP1 and ABI3, it belongs to the B3 domain group. ABI19 is involved in controlling the growth of the embryo and endosperm. Its expression level in both tissues increases

significantly after 8 days. Previous studies have shown that it controls many genes related to seed filling. We found that the promoters of these target genes are open at this time. Many of them contain RY elements, which are known to be bound by ABI3/VP1. This supports the view that ABI19 plays an important role at this stage. Some NAC transcription factors, such as NAC128 and NAC130, also become active. They help manage starch-related genes. We found their binding sites in the open chromatin near starch genes, such as the CACG motif. These NACs are active when starch begins to accumulate, indicating their significant role at this stage.

In the later stage of seed development (about 10 days to maturity) : After the seeds approach the dormant period, the level of abscisic acid (ABA) begins to rise. Meanwhile, many genes related to stress are activated. A key gene at this stage is Viviparous1 (VP1), which is the corn version of ABI3. VP1 is combined with RY components. This helps activate the genes needed in the later stage of embryo development. It can also inhibit genes that may start to germinate prematurely. At this stage, we observed a large number of RY motifs in the open chromatin regions. VP1 was also strongly expressed in our RNA sequencing results. This indicates that VP1 plays an important role as a transcription factor in the later stage of the seed. Other transcription factors also began to play a role, including heat shock factor (HSF) and the bZIP protein similar to ABI5. They help seeds cope with dryness. We found that heat shock elements (AGAAnnTTCT) were present in large quantities in the open areas at this stage. This might imply that the HSF is combined there. Genes such as *ZmHSFA2* and *ABI5* also showed higher expression. These genes may help activate the LEA gene and other protective genes. Chromatin also seems to undergo changes to support dormancy. Transcription factors like ZmABI4 (belonging to the AP2 family) may contribute to this. They can bind to certain open sites and shut down metabolic genes related to germination. These genes remain silent until the seeds absorb water.

In addition to these broad categories, our analysis unearthed TF candidates for more specialized roles or cell-type specific roles. For example, the DOF family TF PBF (P-prolamin-box binding factor), which partners with O2 to regulate zein genes, likely binds accessible sites in late endosperm (we saw its AAAG motif footprints on zein gene promoters). PBF expression is indeed specific to endosperm and peaks during storage phase. Similarly, we observed motifs for MYB and WRKY factors enriched in bundle sheath- or aleurone-specific OCRs (from dissected tissue ATAC-seq), which suggests roles for those TFs in those specific seed tissues. For instance, WRKY TFs may be involved in the aleurone layer's defense gene expression (Yang et al., 2016).

To systematically pinpoint key regulators, we compiled a list of TF genes whose expression profiles matched the pattern of any OCR cluster and whose motif was enriched in those cluster's sequences. This highlighted several candidates as "core" stage-specific TFs (some mentioned above). Many of these have literature support (like O2, ABI3, LEC2), lending credence to the ones that are more novel. For example, we identified a TF of the GRAS family (potentially ZmGRAS11) expressed during endosperm development; GRAS11 is reported to work downstream of O2 in promoting cell expansion, and we indeed saw its motif in mid-stage OCRs. Another novel one was a B3-domain factor apart from ABI3/VP1 – possibly ZmABI19 or ZmLEC1 (though LEC1 is a CCAAT-binding factor, it could show indirectly as a motif). These analyses underscore a temporal division of labor among transcription factors: MADS and ARFs early, various bZIPs and NACs in mid, and B3/ABI3 plus stress TFs late, all orchestrating development in sequence.

4.3 Regulatory network construction and core nodes

One sub-network we constructed focuses on early embryogenesis and endosperm initiation. In this network, a central node is an SVP/AGL MADS-box TF (call it ZmMADSx) which appears to regulate a suite of early genes by binding their promoters (CArG motifs identified in those promoters' OCRs). Those target genes include ones encoding hormone biosynthesis enzymes (e.g. YUCCA for auxin, a GA 20-oxidase for gibberellin) and other TFs like LEC2. Meanwhile, LEC2 itself forms another node that feeds into a slightly later set of targets – we placed LEC2 as regulating genes for seed storage initiation (like LEA precursors) and also upregulating ABI3/VP1 (as suggested by RY motifs upstream of Vp1). Thus, LEC2 is a predicted upstream activator of the maturation network. On the endosperm side, our early network includes NKD1/2 acting on aleurone differentiation genes (e.g. *mrp1*, a transfer layer regulator, and various Betl genes encoding transfer cell proteins). These interactions were

drawn because NKD-binding IDD motifs (TTGTCG) were found in the promoters of *Betl* genes, and expression data shows *Betl* genes drop off in nkd mutants (Zhang et al., 2018).

Moving to mid-development, the network we constructed shows Opaque11 (O11) as a hub in the endosperm regulatory network. O11 (a bHLH) is known to regulate other TFs such as Opaque2, PBF, and DOF3. Our data confirm this hierarchy: we saw O11's binding E-box motif (CANNTG) in the accessible promoters of O2, PBF, and Dof3 during mid-stage, and those genes are indeed co-expressed. Thus in the network, O11 connects to O2, PBF, and DOF3, which in turn connect to their downstream targets (e.g. O2 and PBF jointly to zein storage protein genes, DOF3 and others to sugar metabolism genes). We also included ZmGRAS11 in this mid network, connecting it as a target of O2 (since O2 can transactivate ZmGRAS11) and as a regulator of endosperm cell expansion genes (perhaps cell wall biosynthesis genes). This is consistent with Feng et al. (2018) who identified interactions among these factors. Additionally, ABI19 appears in the mid network bridging embryo and endosperm: ABI19 likely receives input from LEC2 (RY motifs in ABI19 promoter) and outputs to both embryo regulators (like WOX genes for embryo patterning) and endosperm storage genes (it directly activates some late storage genes and coordinates with ABI3/VP1). We depicted ABI19 as a node linking to WOX2 and WUS (key embryo meristem genes), because our ATAC-seq showed ABI19 binding motifs on ZmWOX2 and ZmWUS2 promoters and luciferase assays confirmed ABI19 can bind those promoters. This suggests ABI19 helps set up the embryo body plan while also preparing the seed for maturation.

For the late development network, ABI3/VP1 (the product of Vp1) is the central hub. In our network model, VP1 receives input from earlier factors LEC2 and ABI19 (since both likely help induce Vp1 expression), and then VP1 activates a broad suite of late genes. We connected VP1 to numerous LEA protein genes, late storage genes (e.g. late cysteine-rich peptides), and dormancy-related genes. We also have VP1 inhibiting some targets: known VP1/ABI3-repressed targets include germination-promoting genes like *Amy1* (α-amylase) in cereal aleurone. We saw that the Amy1 promoter has RY motifs and remains inaccessible/low expressed until germination, implying VP1 bound and kept it closed (with the help of VAL repressors). Thus, in the network, VP1 links to Amy1 with a repression arrow. Another piece is the interplay with hormone signaling: VP1 works in concert with ABA. ABA can induce ABI5 (a bZIP) which then binds to late gene ABA-responsive elements. We included ABI5 in the network as being upregulated by VP1 and contributing to activating some late embryogenesis genes. As seeds dry, certain HSFs are also induced by ABA or stress, and our network shows an HSF node connecting to heat-shock protein and chaperone genes in late stage, reflecting their activation.

5 Functional Annotation and Developmental Pathways
5.1 GO and KEGG enrichment of OCR-associated genes
We conducted gene ontology (GO) enrichment and KEGG pathway analysis on gene sets adjacent to open chromatin regions (OCRs), with a focus on genes located near OCRs and whose accessibility changed significantly at different stages of seed development. This helps us identify which biological processes may be affected by changes in chromatin openness. We found that many genes near OCR are associated with seed-related functions. These genes show a strong enrichment in some categories. For instance, GO terms such as "starch biosynthesis process" and "sucrose metabolism process" frequently appear in genes where chromatin accessibility increases in the middle and later stages of seed development. This is consistent with what we already know - the endosperm of corn produces a large amount of starch during this period. Important starch genes, such as *Sh2*, *Bt2* and *Wx*, show strong ATAC-seq signals near their promoters or enhancers in the active state. These open areas may help promote the enrichment of starch-related GO terms. We also witnessed a similar situation in the KEGG analysis. The "starch and sucrose metabolism" pathway was significantly highly expressed in genes with later OCR opening ($p < 0.001$). This indicates that chromatin opening may be involved in activating these genes during the process of storage accumulation.

We also found that many genes near the altered OCR are associated with GO terms related to plant hormones. These terms are mainly related to the synthesis and response of hormones. For instance, "auxin metabolic processes" and "responses to auxin" occur more frequently in genes near early OCR. This makes sense because

auxin plays a crucial role in initiating endosperm growth immediately after fertilization. We discovered open chromatin regions near genes such as *TAR1* (involved in auxin synthesis) and *YUC* (also involved in the same pathway). These open areas suggest that auxin related genes may be activated with the help of nearby accessible DNA in the early stage of seed development.

Similarly, in the later stages of seed development, OCR becomes more accessible near abscisic acid (ABA) signal-related genes, including those encoding PP2C phosphatase and bZIP transcription factors such as ABI5. These changes are consistent with the role of ABA in seed dormancy preparation and the completion of development. In our KEGG analysis, "plant hormone signal transduction" was significantly enriched in genes related to OCR, which supports the view that hormone changes match chromatin openness changes during seed formation.

Another broad category that emerged from enrichment analysis is gene regulation and chromatin organization itself. Several transcription factor activity GO terms (e.g. "sequence-specific DNA binding") and chromatin modification terms (like "histone acetylation") were enriched, meaning that many OCR-linked genes encode regulatory proteins. This includes the key transcription factors we identified (LEC1, O2, VP1, etc.), as well as components of chromatin remodeling complexes. For instance, genes encoding certain SNF2-family chromatin remodelers and chromatin assembly factors became accessible and expressed in specific stages, perhaps to facilitate the large-scale chromatin changes seeds undergo (like endoreduplication in endosperm cells or chromatin desiccation tolerance). Thus, the regulatory architecture appears to be somewhat self-reinforcing – changes in chromatin openness enable the expression of chromatin regulators that further modulate chromatin states downstream (Xie et al., 2023).

5.2 Enrichment of key developmental processes

Focusing on specific developmental processes of interest – such as starch accumulation, cell differentiation, and hormone responses (as mentioned in the outline) – we find clear evidence that these processes are underpinned by coordinated chromatin accessibility changes.

Maize seeds (notably the endosperm) accumulate large amounts of starch, and this process is tightly controlled at the transcriptional level by a suite of enzymes and regulators. Our ATAC-seq data show that many genes in the starch biosynthetic pathway become accessible during the mid-to-late endosperm development. For instance, Sh2 and Bt2, which encode the two subunits of ADP-glucose pyrophosphorylase (the rate-limiting enzyme in starch synthesis), had low chromatin accessibility in early development but showed strong promoter ATAC-seq peaks by 8-10 DAP, concurrent with their increased expression. Similarly, Sugary1 (a starch debranching enzyme) and Waxy (granule-bound starch synthase) genes also gained accessible promoters in the filling stage. GO enrichment already indicated starch biosynthesis was prominent; drilling down, we saw that specific transcription factors known to regulate starch-related genes also had changes in accessibility. ZmNAC128 and ZmNAC130, two NAC family TFs that positively regulate starch accumulation by activating starch synthase genes, showed increased expression in mid-development and had accessible promoter regions, presumably facilitating their own upregulation. It has been demonstrated that mutation in these NAC genes reduces starch content, highlighting their importance. In our data, OCRs containing NAC-binding motifs were found near starch synthase genes, suggesting NACs bind there to boost transcription. In summary, the process of starch accumulation is enriched in the sense that its key players (both enzymes and regulators) are among the genes with significant chromatin opening in the appropriate timeframe.

During seed development, a burst of cell differentiation establishes the various cell types of the embryo (shoot meristem, root meristem, cotyledon tissues) and endosperm (aleurone, starchy cells, transfer cells, etc.). Our findings indicate that genes instrumental in cell differentiation show distinctive chromatin accessibility changes. For example, genes like Leafy Cotyledon1 (Lec1) and Baby Boom, which are involved in embryonic cell fate and somatic embryogenesis, were accessible primarily in early embryogenesis. Lec1 encodes an HAP3 subunit of CCAAT-binding factor and is a master regulator for embryo identity; we found its promoter to be accessible in early embryo stages (consistent with it being active then). In the endosperm, genes that mark specific lineages

such as MRP1 (Myb-related protein 1, specifying transfer cell fate) had a sharp increase in accessibility in the basal endosperm region right after fertilization. MRP1 binds to downstream target gene promoters to initiate transfer cell differentiation; our data show accessible sites in those target promoters (the so-called *BETL* genes) correspondingly present. We noted that gene sets associated with differentiation (like "regulation of meristem development" or "cell fate specification") were enriched in early development clusters of OCR-linked genes. Another example is the set of WOX (WUSCHEL-related homeobox) genes, which control embryo patterning and meristem formation. ZmWOX2 and ZmWOX5 (involved in the apical and root pole of the embryo, respectively) are expressed very early; interestingly, ZmWOX2 was predicted as a target of ABI19 in our network and has promoter RY elements. We observed its promoter is indeed accessible in the early embryo, likely reflecting its activation and importance in the first cell divisions of the zygote. The enrichment of differentiation processes among *OCR* genes underscores that establishing new cell identities – which requires turning on specific sets of genes while turning off others – is accompanied by deliberate remodeling of chromatin at those gene loci. This ensures that lineage-specific genes (e.g. aleurone-specific or embryo shoot-specific) become accessible to transcription machinery only in the correct context.

Plants use different hormones to control the growth of seeds. At first, auxin and cytokinin played a more important role. Afterwards, as the seeds mature and are ready to enter a dormant state, hormones such as abolic acid (ABA) and gibberellin (GA) become even more important. In our research, we found that many genes that respond to auxin - such as those from the Aux/IAA family and the GH3 enzyme - are close to open chromatin spots at an early stage. The promoters of some Aux/IAA genes open just a few days after pollination. This time coincides very well with the early leap in auxin activity during seed development.

The Aux/IAA protein helps control how plants process auxin. They achieve this by working in synergy with ARF (enhanced growth Factor), a protein that helps transmit auxin signals. These obstructions emerged quite early. They may be involved in shaping the way embryos and endosperm form. We found that their promoters opened shortly after pollination. This point in time is in line with the view that auxin takes effect rapidly and uses these repressors to build feedback loops. For abscisic acid (ABA), which becomes more active in the later stage, we studied the *LEA* gene, especially the genes of the Em family. These genes are known targets of ABA and are usually activated by transcription factors such as ABI3 and VP1. At first, their promoter regions did not show too many ATAC-seq signals. But later, these regions opened up, indicating that as the level of abscisic acid rose, these genes were activated. We also found that the GO terms for many genes with late-opening OCR are "response to abscisic acid" and "response to drying". This is in line with the role of abscisic acid in helping seeds dry out and maintain dormancy. A gene related to gibberellin (GA), GA 2-oxidase, also has a late-opening promoter (Du et al., 2023). This gene helps break down gibberellin, which supports the view that an increase in abalic acid levels will reduce gibberellin, thereby preventing early germination.

5.3 Dynamics of pathway activities across development
By tracking the accessibility of chromatin and the changes in gene expression over time, we outline how different biological pathways alter their activity during seed development. In the early stage, from fertilization to a few days after pollination (DAP), genes related to cell growth and basic biosynthetic requirements are particularly active. Our ATAC-seq and RNA-seq data show that the signals of DNA replication and cell cycle genes (including the gene of histone H4) are strong, and histone H4 supports the formation of new chromatin in rapidly dividing nuclei.

At the same time, we noticed brief activity in hormone-related pathways, especially auxin and cytokinin. Several genes induced by auxin showed early transcription and accessible promoters, while type-A cytokinin response regulators also briefly opened up. These patterns match the idea that hormones give a short boost to early seed development, especially for launching endosperm formation. In contrast, ABA-related genes remained mostly inactive in this period—both in terms of expression and chromatin openness—consistent with the low ABA levels before seed maturation begins (Bernardi et al., 2019).

During the middle stage of seed growth (about 5 to 15 days), seeds mainly start to store nutrients. Manufacturing starch has become one of the most important tasks. The promoter regions of most of the key enzymes involved in starch production are open and strongly activated at this stage. In addition to starch, the synthesis of proteins has also begun to accelerate. Genes encoding ribosome parts, translation helper genes and storage proteins (such as zeolin) have become more active. Their promoter regions are also open, which is quite normal as the seeds are busy producing a large amount of protein. To keep all this going, the seeds need more energy. It promotes energy production by activating enzymes involved in glycolysis and the tricarboxylic acid cycle. For instance, the gene promoter of pyruvate kinase is in an open state and has a relatively high expression level. This helps ensure that the seeds have sufficient energy to maintain high levels of starch and protein production.

Transcription factors like Opaque2 and O11 are active here, coordinating these storage processes within the endosperm. Meanwhile, cell division activity drops—mitosis-related genes lose accessibility as endosperm cells shift toward endoreduplication, and embryo development transitions from organ formation to growth and expansion.

In the final phase (from ~20 DAP onward), metabolism slows down, while genes for stress protection, drying, and dormancy take over. We saw reduced chromatin openness at genes like starch synthases, which makes sense as starch accumulation finishes. At the same time, ABA-driven responses peak. LEA proteins, detoxification enzymes, and structural proteins linked to seed hardening show high expression and open promoters.

Antioxidant systems, including peroxidases and the ascorbate-glutathione cycle, are turned on to handle reactive oxygen species during drying. We also saw signs of nutrient recycling, with chromatin opening in genes for proteases and lipases that might break down dying tissues or maternal support cells.

From a control perspective, abscisic acid (ABA) and sugar signaling work together to help seeds enter a dormant state. Genes that help transport sugar usually contain a sugar response part (known as the SURE motif). Most of these genes remain off, possibly due to abscisic acid (ABA). Meanwhile, dormant genes of types like viviparous and DOG1 remain on and retain open DNA regions nearby. This shows how seeds utilize hormone and sugar signals to remain dormant and safe until they grow and mature.

Visually, one can imagine a timeline of seed development where the "activity" of different pathways rises and falls in waves. At first, a developmental wave of growth and patterning (cell cycle, auxin) cresting early, followed by a wave of accumulation (starch/protein biosynthesis) cresting mid-development, and finally a wave of maturation (stress/dormancy) cresting late. Our chromatin accessibility profiles mirror these waves: early developmental pathways exhibit early OCR dynamics, mid pathways have mid-stage OCR peaks, and late pathways have late-stage OCR peaks. These dynamics are, in essence, the epigenomic signature of the maize seed's developmental schedule.

6 Case study: Dynamic Chromatin Accessibility of Specific Genes at Key Stages
6.1 Chromatin accessibility changes at key endosperm development genes
The endosperm of maize differentiates into multiple cell types, with the outer aleurone layer and inner starchy endosperm being the primary ones. Proper formation of the aleurone is essential for seed nutrient mobilization during germination and overall seed viability. A master regulator of aleurone cell fate in maize is the transcription factor Naked Endosperm 1 (NKD1), along with its duplicate NKD2. These are INDETERMINATE DOMAIN (IDD) family transcription factors required to specify aleurone versus starchy endosperm identity; nkd mutants develop extra layers of starchy cells in place of aleurone, hence the "naked endosperm" phenotype (Figure 2) (Hughes et al., 2023). We examined the ATAC-seq and RNA-seq profiles for the *NKD1* gene and some of its known targets during seed development.

Early in seed development (around 3–4 DAP), we found that the promoter region of *NKD1* itself became highly accessible. This timing corresponds to endosperm cellularization and the onset of aleurone differentiation at the kernel periphery. The *NKD1* gene's promoter showed a clear ATAC-seq peak specifically in the peripheral

endosperm tissue (when we analyzed dissected nucellus/aleurone vs. central endosperm), indicating that chromatin opening at *NKD1* is localized to the future aleurone layer. By 6 DAP, *NKD1* transcripts were abundant in the RNA-seq, and its promoter remained accessible, suggesting *NKD1* is actively expressed through the differentiation phase. Notably, we identified an OCR not only at the core promoter but also in an intronic region of *NKD1* that contains an ID1-binding motif; this could be an autoregulatory element or a site for feedback regulation by NKD itself or its dimerization partners. The accessibility of these *NKD1* regulatory regions is consistent with the gene's functional activation precisely when aleurone cells are being specified.
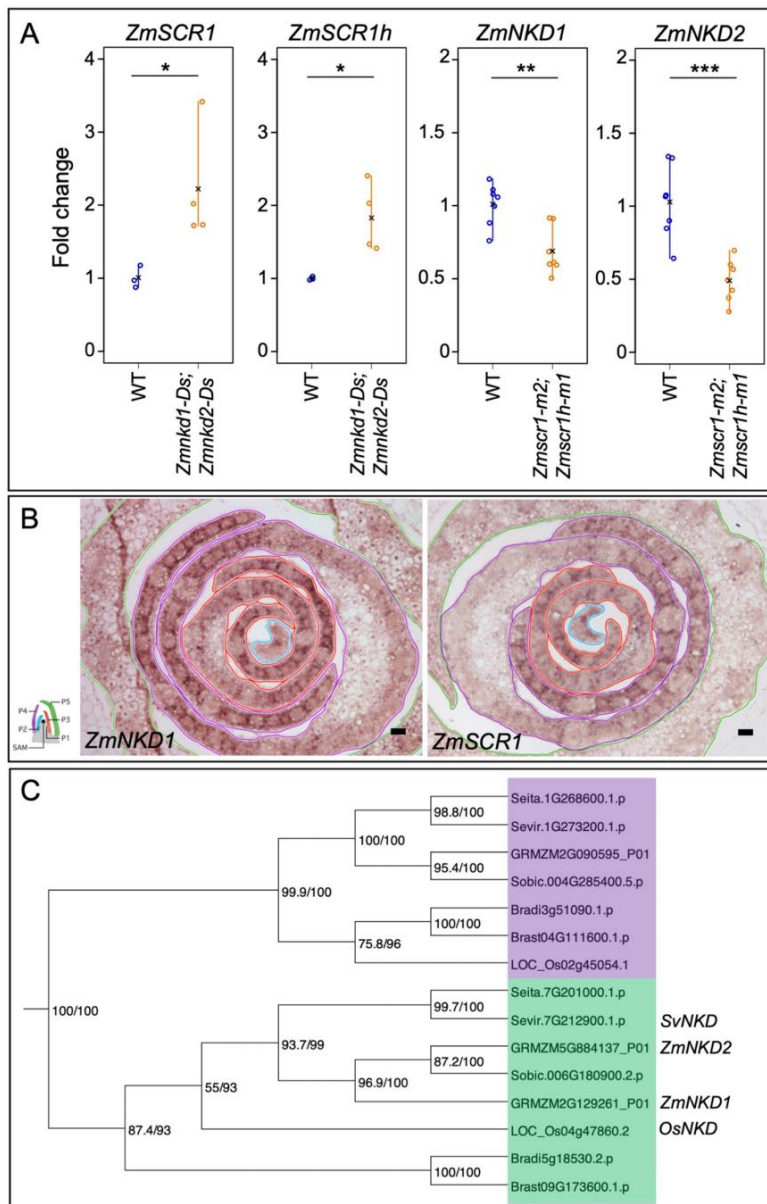

Figure 2 *NKD* and *SCR* transcripts accumulate in the same spatial domain with levels determined by a feedback loop (Adopted from Hughes et al., 2023)

Image caption: A) Quantitative RT-PCR of *ZmSCR1* and *ZmSCR2* transcripts in the *Zmnkd1-Ds;Zmnkd2-Ds* mutant, and *ZmNKD1* and *ZmNKD2* in the *Zmscr1-m2;Zmscr1h-m2* mutant. Open circles are individual biological replicates, black crosses indicate the mean for each genotype. Statistical significance as calculated on log-transformed fold change data by Welch's t-test (two-tailed) indicated above each plot: *$P\leq0.05$; **$P\leq0.01$; ***$P\leq0.001$. B) *In-situ* hybridization to *ZmNKD1* and *ZmSCR1* in maize wild-type B73 apices. In each image the P2 primordium is outlined in blue, P3 in red, P4 in purple and P5 in green, as indicated in the adjacent cartoon diagram. Darker purple signal represents successful hybridization to each transcript of interest. Scale bars: 50 μ m. C) Maximum likelihood phylogeny of the *NKD* genes in monocots. The *NKD* clade is highlighted in green, and the adjacent monocot clade in purple. Bootstrap values are displayed at each branch of the phylogeny (Adopted from Hughes et al., 2023)

We next looked at *NKD1* target genes. One direct target is the gene *MRP1* (Myb-related protein-1), a transcription factor that *NKD1* activates to promote transfer cell (basal endosperm transfer layer, BETL) differentiation. In our data, *MRP1* showed a distinct open chromatin profile: its promoter was accessible slightly later than *NKD1* – around 4–6 DAP – aligning with the formation of the BETL after the aleurone begins to differentiate. *MRP1* expression and accessibility were localized to the basal endosperm region, as expected. Interestingly, *MRP1*'s promoter has multiple IDD (NKD-binding) motifs, and we saw footprints on those motifs in the ATAC-seq data from basal endosperm nuclei, strongly suggesting NKD1/2 proteins bind there to activate *MRP1*. Furthermore, genes that *MRP1* in turn regulates (the so-called *BETL* genes encoding transport proteins like BETL1, BETL2, etc.) also showed stage- and tissue-specific chromatin accessibility: their promoters opened and transcripts accumulated in the basal endosperm slightly after MRP1 was expressed. For example, the promoter of Betl1 (a defensin-like transport protein) had no ATAC signal at 3 DAP, became accessible at 6 DAP specifically in basal endosperm tissue, and was highly accessible by 8 DAP – matching the Betl1 expression pattern. This cascade – *NKD1* accessible first, then *MRP1*, then *BETL* genes – paints a coherent picture of the gene regulatory cascade in aleurone/BETL differentiation, each step evident in our chromatin data.

One of the important functions of *NKD1* is to ensure that only one layer of batter is formed on the outside. *NKD1* and *NKD2* achieve this by inhibiting a gene called *Thick aleurone layer 1* (*Thk1*). If *Thk1* remains active, it may result in more than one layer of paste powder. We examined the area around the *Thk1* gene. In seeds that simultaneously lack both *NKD1* and *NKD2*, the chromatin near *Thk1* is in an open state. This result is consistent with the research findings of Gontarek et al. (2016). However, in normal seeds with *NKD1* present, the same area remains closed. This indicates that *NKD1* may introduce other proteins to help *Thk1* remain closed. It may also block the gene in a roundabout way. We also discovered a possible silent area upstream of *Thk1*. This region contains IDD motifs and is more closed in normal seeds than in mutants. However, we lack a large amount of mutant data, so this part of the research still needs further study. Despite this, *NKD1* may act in synergy with repressor proteins at certain sites. This is in line with another point we saw before - open chromatin does not always mean that genes are activated. Sometimes, it allows repressor proteins to enter and shut down genes. In short, the case of *NKD1* shows that observing the openness of chromatin can explain how seed tissues such as Aladdin and BETL are formed. When *NKD1* is activated, it triggers a chain reaction of gene activity. We can achieve this by tracking the time when the open region of each gene locus appears.

6.2 Candidate regulatory elements linked to seed maturation traits
One powerful application of our chromatin accessibility map is the identification of regulatory DNA associated with important seed traits. Many agronomic traits, such as seed size, composition (starch/protein/oil content), and stress tolerance of seeds, are quantitative and controlled by multiple loci. Genome-wide association studies (GWAS) in maize have identified numerous trait-associated single nucleotide polymorphisms (SNPs) that often lie in noncoding regions. A recurring challenge is to pinpoint which SNPs affect gene regulation and how. Our ATAC-seq data help address this by showing which noncoding regions are likely functional (being open chromatin) and therefore which trait-associated SNPs might reside in bona fide regulatory elements.

As part of this study, we cross-referenced a list of known seed trait QTL and GWAS hits with our OCR locations. We found that a substantial proportion of trait-associated SNPs for seed traits are indeed located within or very close to OCRs in our dataset (far more than would be expected by chance, which is consistent with other findings that phenotype-associated variants often overlap accessible chromatin). For example, a known QTL for kernel weight maps near the gene *ZmABI19*. Our data show an OCR (enhancer) in the 5′ region of *ZmABI19* that is highly accessible in mid-development. Interestingly, within this enhancer resides a SNP that had been associated with variation in kernel size in a diversity panel. The SNP allele correlated with larger kernels is predicted to create a stronger binding site for an ABI3/VP1 protein (changing the sequence closer to the RY consensus). The chromatin at *ZmABI19*'s enhancer was accessible regardless of SNP variant, but the downstream effect might be that one allele of this OCR drives higher ABI19 expression. This illustrates how an open chromatin region harboring a nucleotide polymorphism can contribute to phenotypic diversity – a stronger enhancer could lead to

more ABI19, potentially prolonging grain filling and increasing seed size (consistent with the allele effect). This candidate regulatory element now merits functional testing (e.g. reporter assays or genome editing to swap alleles).

Another striking example involves the NKD1 locus and a trait not immediately obvious: leaf length. A GWAS had identified a SNP in the promoter of NKD1 that was associated with variation in maize leaf length (a vegetative trait). This SNP lies within an OCR (the NKD1 promoter is accessible during seed development as we discussed) (Zhu et al., 2024). One hypothesis is that the NKD1 allele that affects leaf length does so indirectly via seed development – possibly by altering aleurone function or hormone balance in the seed that has carry-over effects on seedling vigor and leaf growth. To test the effect of this regulatory SNP, we used CRISPR-Cas9 genome editing to create small deletions in the NKD1 promoter that encompassed the SNP site (essentially generating NKD1^{pro-1} and NKD1^{pro-2} in our study). These edited lines had subtle but clear phenotypic differences: their third and second leaves (14 days after germination) were significantly longer than those of wild-type plants. This phenocopies the association of the particular SNP allele (longer leaves), thereby confirming that the regulatory region containing that SNP influences leaf development. Our ATAC-seq data showed that the edited region is normally an active promoter OCR in seeds; altering it likely changed NKD1 expression in the seed (though NKD1 is mostly seed-specific, so the mechanistic connection to leaves might be via altered seed nutrient profile or hormone provisioning). Nonetheless, this result demonstrates that manipulating an open chromatin region identified in seed development can have measurable trait outcomes, underscoring the value of open chromatin maps for pinpointing such elements.

We also identified candidate regulatory elements for seed composition traits. For example, oil content in maize kernels is largely determined by the embryo (which is rich in lipids). A major QTL for oil content is located on chromosome 6 (the qHO6 locus). Our chromatin map revealed an enhancer region within that QTL interval that becomes accessible specifically in the embryo during mid-development. This enhancer is near a gene encoding an AP2-domain transcription factor that is expressed in the embryo and known to affect lipid accumulation (potentially an ortholog of WRI1, an AP2 factor controlling oil biosynthesis in Arabidopsis). The presence of an embryo-specific OCR at this locus suggests this could be the causal regulatory region for the oil QTL; SNPs altering this enhancer's strength might lead to differences in expression of the AP2 factor and subsequently lipid biosynthesis genes. This hypothesis is reinforced by the motif content – the enhancer contained multiple RY motifs (which AP2/VRN1 factors might be part of a complex with ABI3), hinting at complex regulation. We consider this enhancer a top candidate for fine-mapping and editing to boost oil content (Zhou et al., 2024b).

6.3 Epigenetic accessibility as a guide for genome editing

The knowledge we learn from open chromatin can truly assist in crop breeding, especially when using tools such as CRISPR. ATAC-seq can show which parts of DNA are active. These active sites are ideal locations for editing, which may change the way genes work or the traits of plants. If we do not want to change the gene itself, we can edit nearby open regions, such as promoters or enhancers. These regions control the frequency of gene usage. Our research results support the concept of "cis engineering". This means using CRISPR/Cas to adjust these control sites, thereby altering the behavior of genes without touching the genetic code.

One of the clearest demonstrations in our study is the NKD1 promoter editing discussed above. By deleting a small region of an accessible promoter containing a trait-associated SNP, we were able to mimic the effect of a natural variant and confirm its phenotypic influence. Notably, the edited plants did not show obvious deleterious effects on seed development or plant viability, suggesting that fine-tuning regulatory regions can produce a subtle improvement in a trait (leaf growth in this case) without negative trade-offs. This is a promising outcome for breeding – it implies that targeting regulatory alleles (as opposed to complete knockouts of genes) might yield beneficial phenotypic variation that is less likely to harm overall fitness, because it's making quantitative rather than qualitative changes in gene expression.

Another scenario in which open chromatin data guide editing is when creating synthetic promoters or enhancers. If a certain pathway needs to be upregulated, one might want to insert an enhancer or strengthen an existing one.

Knowing which enhancers are naturally active in seeds (and under what conditions) helps choose candidates for enhancement. For instance, we identified an enhancer upstream of a key starch synthesis gene that is moderately active in mid-endosperm. If the goal is to increase starch content, one strategy could be to engineer that enhancer to have additional TF binding sites (e.g., more copies of O2 binding site) to drive higher expression of the starch gene. Since we see that enhancer is accessible, adding binding motifs for endosperm-active TFs (like O2, PBF) might boost its activity. Conversely, to reduce the expression of a gene (perhaps to redirect metabolic flow), one could introduce small mutations in its enhancer motifs to weaken TF binding (effectively what some natural SNPs do). Because our data pinpoint the exact motifs used by the plant (footprints), genome editing can be extremely precise - altering one or two base pairs in a motif can disrupt a TF's binding. This is much cleaner than traditional methods (like RNAi or random mutagenesis) because it leaves the rest of the genome untouched and fine-tunes only the target gene's regulatory control.

We also emphasize that open chromatin regions are potential sites where epigenetic modifications could be applied. For example, technologies are emerging to deposit DNA methylation or repressive marks at specific sequences to downregulate gene activity (epigenome editing). The accessible regions we identified would be logical targets for such epigenetic editing, because adding a repressive mark to an enhancer or promoter that is normally open could shut down the associated gene. This could be reversible and potentially heritable, offering a new breeding tool that doesn't alter the DNA code but the chromatin state. In crops, one might imagine using epigenome editing to temporarily silence a gene that limits yield, and our map tells us where to target such modifications (e.g., the promoter of a negative regulator of grain filling).

Editing noncoding regulatory sequences can reproduce quantitative trait variation. Hendelman et al. (2021) in tomato, for instance, dissected a homeobox gene's enhancer and created alleles that mimic natural diversity in fruit size. Similarly, in maize, a study by Wang et al. (2021) edited a promoter in a stem cell regulatory circuit to demonstrate how cis variation influences meristem architecture. We contribute to this growing evidence by showing an example in maize seed (NKD1) and by highlighting numerous new candidates in seed development that could be edited for improvement. The union of ATAC-seq and CRISPR thus holds great promise: ATAC-seq flags the critical regulatory nodes, and CRISPR allows us to modulate them in targeted ways.

In practical breeding programs, this approach could be used to produce elite alleles faster. If a favorable allele is known (through GWAS) but is in a linkage drag context, one could recreate it by editing the native variety's regulatory sequence as we did. For traits like seed size, composition, or stress tolerance, where often the differences are due to promoter or enhancer variants, genome editing guided by our open chromatin map could accelerate the incorporation of those traits. Another advantage is reducing pleiotropy: by tweaking a gene's expression rather than knocking it out, one might avoid undesirable effects in other tissues or stages. For instance, completely knocking out NKD1 or ABI3 causes severe defects (nkd leads to defective aleurone, abi3 mutants cause precocious germination), but a slight promoter variant might improve one aspect (leaf growth or stress tolerance) without fully losing function. Indeed, our NKD1[pro] edited lines were perfectly viable and just had subtly longer leaves, implying the core seed functions of NKD1 remained intact.

7 Discussion and Outlook

The results of this study indicate that chromatin accessibility does not merely follow genetic activity - it plays a crucial role in shaping seed traits. By turning on or off certain DNA regions at the right time, it controls when and where genes are activated. These changes will affect important traits, such as the size, strength of the seeds and the storage capacity of nutrients. For instance, at the beginning of the grain filling period, we found that the promoter of the starch production gene became more open. This openness enables these genes to be activated and drive the accumulation of starch, thereby controlling the carbohydrate content that eventually enters the seed. If the chromatin around these genes remains closed, the genes will not be activated even with the correct signals. Therefore, chromatin opening is a necessary condition for high starch production. We also found that the formation of a single aleurone layer in the endosperm requires the activation of NKD1 at the correct time. This occurs when its regulatory DNA becomes more open. If the chromatin does not undergo such changes due to

mutations or stress, the aleurone layer may not form correctly, which could damage the seed's ability to store and utilize nutrients. This tells us that these traits depend on timely gene activity, and the openness of chromatin is part of the control system, which ensures that everything occurs in the right sequence and position.

Data further show that many differences in seed traits are tied to regulatory DNA, not the gene sequences themselves. We saw that SNPs linked to traits often sit in open chromatin areas. This means that both evolution and breeding may have worked on these regions, picking versions of DNA that keep chromatin more or less open at key spots, which affects how traits show up. For breeders, this is a useful insight. It's a reminder to not just look at gene mutations, but also at nearby DNA that controls when those genes turn on. Even small changes in gene activity—driven by chromatin or nearby DNA elements—can cause big trait differences, like oil level or seed size. In practice, knowing how chromatin affects traits might make breeding more accurate. For example, adding markers from open chromatin regions to breeding models could help pick better plants. These markers likely matter because they control how genes behave.

Single-cell ATAC-seq is another frontier that stands to revolutionize our understanding of seed development. The maize seed contains multiple tissues and cell types (embryo shoot, root, scutellum, various endosperm cell types, etc.), each with unique regulatory profiles. Our data, though high-resolution in time, was generally from whole-seed or bulk endosperm tissues. Single-cell ATAC-seq would allow us to parse out cell-type-specific chromatin landscapes. In a complex tissue like endosperm, this is particularly valuable – e.g., we could recover separate ATAC profiles for aleurone vs. starchy cells vs. BETL cells from a single assay. This would disentangle composite signals and possibly identify cell-type-specific enhancers that were diluted in bulk data. Encouragingly, techniques for single-cell ATAC in plants (including maize) are emerging. Coupling those with single-cell RNA-seq (which has already been applied to maize seeds for cell-type networks) would yield a powerful multi-omic atlas: for each cell type, know its chromatin accessibility and gene expression and how they change over development. Such a resource would immensely clarify how different seed compartments coordinate (for example, how the embryo "talks" to endosperm via chromatin changes and signaling molecules).

The combined use of multiple omics tools can accelerate gene discovery and facilitate crop research. In the past, searching for the genes behind QTL (quantitative trait loci) was time-consuming and laborious, and it was difficult to determine which part of the genome was crucial. But now, with the help of methods such as ATAC-seq and other omics data, it is easier to identify useful genes or control regions. In our research, we discovered an enhancer within the QTL that controls the oil content. This region is associated with an AP2 transcription factor. As we collect more datasets - from different growth periods, stress environments and plant types - we can start to build large genetic networks. Machine learning can help complete this step. Tools like WGCNA can group genes with similar expressions. Similarly, ATAC-seq can identify DNA regions that are open or closed to each other. These patterns might also reveal the three-dimensional folding patterns of the genome, especially when folding occurs under stress. So far, such comprehensive data have been used to construct the genetic maps of Arabidopsis roots and corn leaves. If we apply the same method to seeds, it may help us understand seed dormancy and how seeds store nutrients.

Acknowledgments

Conflict of Interest Disclosure

The author affirms that this research was conducted without any commercial or financial relationships that could be construed as a potential conflict of interest.

References

Bernardi J., Battaglia R., Bagnaresi P., Lucini L., and Marocco A., 2019, Transcriptomic and metabolomic analysis of ZmYUC1 mutant reveals the role of auxin during early endosperm formation in maize, Plant Science, 281: 133-145.
https://doi.org/10.1016/j.plantsci.2019.01.027

Bubb K., Hamm M.O., Tullius T.W., Min J.K., Ramirez-Corona B., Mueth N.A., Ranchalis J., Mao Y.Z., Bergstrom E.J., Vollger M.R., Trapnell C., Cuperus J.T., Stergachis A.B., Queitsch C., and Stergachis A., 2024, The regulatory potential of transposable elements in maize, bioRxiv, 602892: 1-33.
https://doi.org/10.1101/2024.07.10.602892

Du K., Zhao W., Lv Z., Liu L., Ali S., Chen B., Hu W., Zhou Z., and Wang Y., 2023, Auxin and abscisic acid play important roles in promoting glucose metabolism of reactivated young kernels of maize (*Zea mays* L.), Physiologia Plantarum, 175(5): e14019.

https://doi.org/10.1111/ppl.14019

Du Y., Liu L., Peng Y., Li M., Li Y., Liu D., Li X., and Zhang Z., 2020, *UNBRANCHED3* expression and inflorescence development is mediated by *UNBRANCHED2* and the distal enhancer, *KRN4*, in maize, PLoS Genetics, 16(4): e1008764.

https://doi.org/10.1371/journal.pgen.1008764

Feng F., Qi W., Lv Y., Yan S., Xu L., Yang J., Yuan Y., Chen Y., Zhao H., and Song R., 2018, OPAQUE11 is a central hub of the regulatory network for maize endosperm development and nutrient metabolism, The Plant Cell, 30(2): 375-396.

https://doi.org/10.1105/tpc.17.00616

Galli M., Khakhar A., Lu Z., Chen Z., Sen S., Joshi T., Nemhauser J., Schmitz R.J., and Gallavotti A., 2018, The DNA binding landscape of the maize AUXIN RESPONSE FACTOR family, Nature Communications, 9: 4526.

https://doi.org/10.1038/s41467-018-06977-6

Gontarek B.C., Neelakandan A., Wu H., and Becraft P., 2016, NKD transcription factors are central regulators of maize endosperm development, Plant Cell, 28(11): 2916-2936.

https://doi.org/10.1105/tpc.16.00609

Hendelman A., Zebell S., Rodriguez-Leal D., Dukler N., Robitaille G., Wu X., Kostyun J., Tal L., Wang P., Bartlett M., Eshed Y., Efroni I., amd Lippman Z.B., 2021, Conserved pleiotropy of an ancient plant homeobox gene uncovered by cis-regulatory dissection, Cell, 184(6): 1724-1739.

https://doi.org/10.1016/j.cell.2021.02.001

Hughes T.E., Sedelnikova O., Thomas M., and Langdale J., 2023, Mutations in *NAKED-ENDOSPERM* IDD genes reveal functional interactions with *SCARECROW* during leaf patterning in C_4 grasses, PLOS Genetics, 19(1): e1010715.

https://doi.org/10.1371/journal.pgen.1010715

Lu Z., Hofmeister B.T., Vollmers C., DuBois R., and Schmitz R.J., 2016, Combining ATAC-seq with nuclei sorting for discovery of cis-regulatory regions in plant genomes, Nucleic Acids Research, 45: e41.

https://doi.org/10.1093/nar/gkw1179

Wang B., Aguirre L., Rodríguez-Leal D., Hendelman A., Benoit M., and Lippman Z.B., 2021, Dissecting cis-regulatory control of quantitative trait variation in a plant stem-cell circuit, Nature Plants, 7(4): 419-427.

https://doi.org/10.1038/s41477-021-00898-x

Xie S., Tian R., Zhang J., Liu H., Li Y., Hu Y., Yu G., Huang Y., and Liu Y., 2023, *Dek219* encodes DICER-LIKE1 protein that affects chromatin accessibility and kernel development in maize, Journal of Integrative Agriculture, 22(10): 2961-2980.

https://doi.org/10.1016/j.jia.2023.02.024

Yang J., Ji C., and Wu Y., 2016, Divergent transactivation of maize storage protein zein genes by the transcription factors Opaque2 and OHPs, Genetics, 204(2): 581-591.

https://doi.org/10.1534/genetics.116.192385

Yang T., Wang H., Guo L., Wu X., Xiao Q., Wang J., Wang Q., Ma G., Wang W., and Wu Y., 2022, ABA-induced phosphorylation of basic Leucine Zipper 29, ABSCISIC ACID INSENSITIVE 19 and Opaque2 by SnRK2.2 enhances gene transactivation for endosperm filling in maize, The Plant Cell, 34(5): 1933-1956,.

https://doi.org/10.1093/plcell/koac044

Yu G., Shoaib N., Yang Y., Liu L., Mughal N., Mou Y., and Huang Y., 2023, Effect of phosphorylation sites mutations on the subcellular localization and activity of AGPase Bt2 subunit: implications for improved starch biosynthesis in maize, Agronomy, 13(8): 2119.

https://doi.org/10.3390/agronomy13082119

Yuan L., Song X., Zhang L.S., Yu Y., Liang Z., Lei Y., Ruan J., Tan B., Liu J., and Li C., 2020, The transcriptional repressors VAL1 and VAL2 recruit PRC2 for genome-wide Polycomb silencing in *Arabidopsis*, Nucleic Acids Research, 49(1): 98-113.

https://doi.org/10.1093/nar/gkaa1129

Zhan J., Li G., Ryu C.H., Ma C., Zhang S., Lloyd A., Hunter B.G., Larkins B.A., Drews G.N., Wang X.F., and Yadegari R., 2018, Opaque-2 regulates a complex gene network associated with cell differentiation and storage functions of maize endosperm, Plant Cell, 30(10): 2425-2446.

https://doi.org/10.1105/tpc.18.00392

Zhang X., Yang Y., Liang W., and Zhang D., 2018, The MADS-box gene family in maize: genome-wide characterization and expression analysis during reproductive development and abiotic stress responses, Frontiers in Plant Science, 9: 1281.

https://doi.org/10.3389/fpls.2018.01281

Zhang Z., Lin L., Chen H., Ye W., Dong S., Zheng X., and Wang Y., 2022, ATAC-seq reveals the landscape of open chromatin and *cis*-regulatory elements in the *Phytophthora sojae* genome, Molecular Plant-Microbe Interactions, 35(4): 301-310.

https://doi.org/10.1094/MPMI-11-21-0291-TA

Zheng G., Wu J., Li J., Zhao Y., Zhou C., Ren R., Wei Y., Zhang X., and Zhao X., 2025, The chromatin accessibility landscape during early maize seed development, The Plant Journal, 121(6): e70073.

https://doi.org/10.1111/tpj.70073

Zhou L., Bao Y., Wang J.E., Wang S.L., Zhong W.X., and Sun X.R., 2024a, Nucleotide polymorphism in *Zea*: patterns and influences on crop traits, Molecular Plant Breeding, 15(5): 220-232.

https://doi.org/10.5376/mpb.2024.15.0022

Zhou L., Bao Y., Zhang D.N., Wang S.L., Zhang Y.J., and Yu X.T., 2024b, QTL mapping of resistance to ear rot in maize based on SNP markers and improvement of high-yield and disease-resistance traits, Molecular Plant Breeding, 15(6): 340-350.

https://doi.org/10.5376/mpb.2024.15.0032

Zhu Y., Ngan H., Zhu T., Nan L., Li W., Xiao Y., Zhuo L., Chen D., Tu X., Gao K., Yan J., Zhong S., and Yang N., 2024, Pan-cistrome analysis of the leaf accessible chromatin regions of 214 maize inbred lines, bioRxiv, 618191: 1-32.

https://doi.org/10.1101/2024.10.14.618191