

# SCIENTIFIC REPORTS



OPEN

## Spatio-temporal analysis of coding and long noncoding transcripts during maize endosperm development

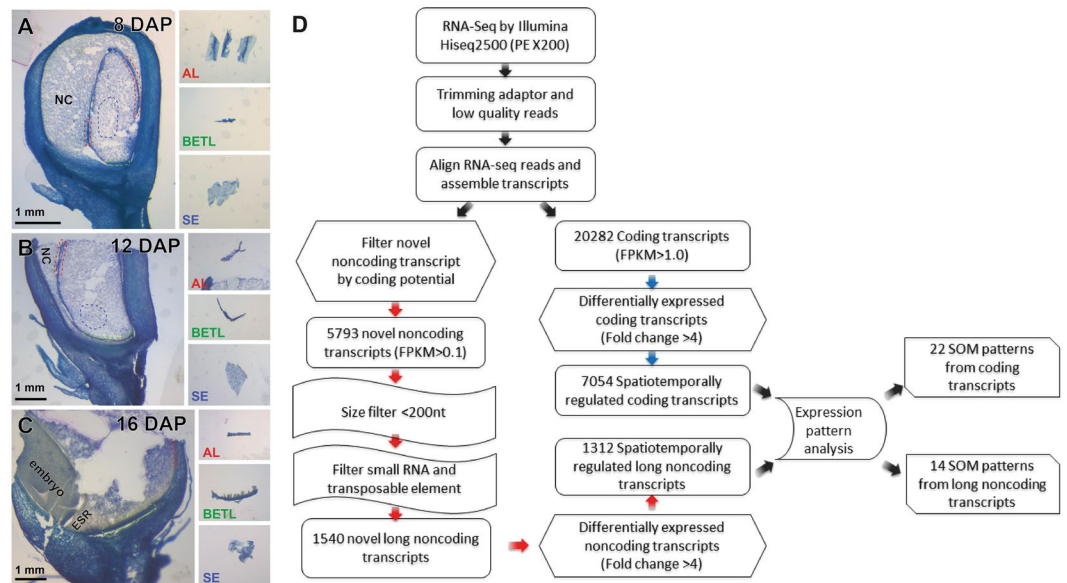
Eun-Deok Kim<sup>1</sup>, Yuqing Xiong<sup>2</sup>, Youngjae Pyo<sup>1</sup>, Dong-Hwan Kim<sup>1</sup>, Byung-Ho Kang<sup>2,3</sup> & Sibum Sung<sup>1</sup>

The maize endosperm consists of three major compartmentalized cell types: the starchy endosperm (SE), the basal endosperm transfer cell layer (BETL), and the aleurone cell layer (AL). Differential genetic programs are activated in each cell type to construct functionally and structurally distinct cells. To compare gene expression patterns involved in maize endosperm cell differentiation, we isolated transcripts from cryo-dissected endosperm specimens enriched with BETL, AL, or SE at 8, 12, and 16 days after pollination (DAP). We performed transcriptome profiling of coding and long noncoding transcripts in the three cell types during differentiation and identified clusters of the transcripts exhibiting spatio-temporal specificities. Our analysis uncovered that the BETL at 12 DAP undergoes the most dynamic transcriptional regulation for both coding and long noncoding transcripts. In addition, our transcriptome analysis revealed spatio-temporal regulatory networks of transcription factors, imprinted genes, and loci marked with histone H3 trimethylated at lysine 27. Our study suggests that various regulatory mechanisms contribute to the genetic networks specific to the functions and structures of the cell types of the endosperm.

Maize (*Zea mays* L.) is a major source of carbohydrate, protein, fat, and other nutrients for humans and livestock through its grain and biomass<sup>1–4</sup>. Maize endosperm development after fertilization involves unique and well-defined spatio-temporal cellular differentiation events<sup>5,6</sup>. Three distinct tissue types of the maize endosperm are the starchy endosperm (SE), the basal endosperm transfer cell layer (BETL), and the aleurone cell layer (AL)<sup>2–4,7,8</sup>. The SE makes up the bulk of the endosperm and consists of storage cells packed with starch and storage proteins<sup>9,10</sup>. The BETL and AL are at the endosperm surface, and these cells form an epidermis-like layer that encloses the entire endosperm and embryo. The BETL is adjacent to the maternal pedicel tissue, serving as the main interface between the maternal and filial tissues<sup>11,12</sup>. The BETL functions to mediate maternal-to-filial transport and defence against microbial attack. The primary function of the AL is to mobilize nutrients stored in the endosperm during seed germination<sup>7,8</sup>. Distinct functions of the three cell types in the endosperm make it an excellent model for study of cell differentiation and gene regulation mechanisms. Also, the availability of the maize genome sequence provides an opportunity to combine the existing rich genetic resources with genomics approaches<sup>13</sup>.

Stable gene expression in differentiated cells is often maintained by epigenetic regulation that involves chromatin modifications. Epigenetic changes in the gene regulatory network contribute to proper differentiation and maintenance of cell identity. Given that BETL, AL, and SE are derived from a single cell type but that they express distinct sets of genes, studying chromatin modifications involved in their specialization could reveal molecular mechanisms of epigenetic regulation controlling cell differentiation in eukaryotes. In addition, there is growing awareness that eukaryotic cells produce many long noncoding RNAs (lncRNAs) that play roles in various developmental programs. Mis-regulation of lncRNAs can lead to developmental defects, suggesting their contribution

<sup>1</sup>Department of Molecular Biosciences and Institute for Cellular and Molecular Biology, University of Texas at Austin, Austin, TX, 78712, USA. <sup>2</sup>Microbiology and Cell Science, University of Florida, Gainesville, FL, 32611, USA. <sup>3</sup>School of Life Sciences, State Key Laboratory for Agrobiotechnology, The Chinese University of Hong Kong, Shatin, New Territories, Hong Kong, China. Correspondence and requests for materials should be addressed to B.-H.K. (email: [bkang@cuhk.edu.hk](mailto:bkang@cuhk.edu.hk)) or S.S. (email: [sbsung@austin.utexas.edu](mailto:sbsung@austin.utexas.edu))



**Figure 1.** Schematic of method for transcriptome analysis of developing maize endosperm. (A–C) HistoGene-stained frozen sections from kernels at (A) 8 DAP, (B) 12 DAP, and (C) 16 DAP. Three insets on the right side of each panel show the three tissue types isolated from cryo-sections by free-hand dissection. The tissues surrounded by the red dashed lines and the blue dashed lines correspond to the AL and SE tissue, respectively. The BETL is marked with green dashed lines. NC, nucellus. (D) The bioinformatics pipeline used to identify spatio-temporally regulated coding and long noncoding transcripts at three stages of the maize endosperm development.

to the regulation of gene expression during development<sup>14</sup>. lncRNAs are essential *cis* and *trans* regulators of gene activity that can function as scaffolds for chromatin-modifying complexes and precursors for regulatory RNAs. Although several lncRNAs have been known for decades, it has only been appreciated recently with the advent of high-throughput RNA sequencing (RNA-Seq) that approximately 10- to 20- fold more genomic sequence is transcribed into noncoding RNA than into protein-coding RNA<sup>15</sup>. Despite growing numbers of identified lncRNAs in eukaryotes, functions and underlying mechanisms of lncRNA action are still poorly understood.

Several previous studies analysed transcriptomes during maize endosperm development<sup>16–26</sup>. These studies focused on transcripts obtained from either dissected endosperm tissues at a certain developmental stage<sup>10, 16, 18, 19, 22, 23, 25, 26</sup> or whole endosperms from different developmental stages<sup>17, 20, 21, 24</sup>. In this study, we identified transcripts that are differentially regulated during the endosperm development by carrying out a spatio-temporal analysis of coding and noncoding transcripts in the three endosperm cell types at three stages of their development. Our analysis indicates that the transcriptomic landscapes of BETL, AL, and SE are distinct and, within the cells of the BETL, temporal expression profiles swing more significantly than in the other cells of the endosperm. Correlative changes in histone modifications and expression of lncRNAs in the vicinity of genes exhibiting spatio-temporal patterns suggest that gene regulatory programs involving chromatin modification operate in cell-type-specific manners in the maize endosperm during its development.

## Results

**Cell-type-specific transcriptome analysis during endosperm development.** To address spatio-temporal transcriptional profiles over the course of endosperm development, we analysed transcriptomes of BETL, AL, and SE tissues at three different developmental stages. At 8, 12, and 16 days after pollination (DAP), fresh kernels were harvested and frozen in the field with liquid nitrogen. Because cells in the three tissues are spatially segregated in the endosperm, we were able to isolate cell specimens of BETL, AL, and SE from the cryo-sections (Fig. 1A–C). The embryo-surrounding region (ESR) is another endosperm tissue, but we did not include ESR cells in our analysis because ESR differentiation is tightly associated with the embryo development<sup>3</sup>. We prevented cross-contamination from ESR by isolating specimens from at least 2 mm away from the embryo. The ESR tissue is readily distinguished from BETL and SE in cryo-sections (Fig. 1C).

Poly (A)-enriched mRNA samples isolated from each cell type at each of the three different developmental stages were used to construct nine sequencing libraries. Paired-end Illumina sequencing generated an average of 20 million reads of 200-nucleotide sequence tags for subsequent RNA-Seq analysis (Table S1). The sequencing reads were aligned to the maize genome using ZmB73\_RefGen\_v2<sup>13</sup> by Tophat<sup>27</sup>. More than 80% of sequence reads were mapped to the reference genome (Table S1). The aligned reads were then reassembled using the Cufflinks assembly<sup>28</sup> with ZmB73\_5b\_filtered set, and fragments per kilobase of exon per million fragments mapped (FPKM) values were calculated for further analysis. The coding transcripts with FPKM > 1.0 and the noncoding transcripts with FPKM value > 0.1 were used for further analysis (Fig. 1D).

We first examined the annotated transcript population. Coding transcripts in the combined RNA-Seq data sets represented a total of 20,282 (out of the total 39,656) annotated genes, although the total number of genes identified in each tissue type at the three developmental stages was variable (Fig. S1A). Across the three different developmental stages, the higher numbers of coding transcripts were observed in BETL samples than in the other two cell types (BETL, 17,288; AL, 16,878; and SE, 15,570; Fig. S1A). It has been speculated that transcriptional regulation in BETL and AL is most active between 8 DAP to 16 DAP due to the remarkable cell differentiation in the tissues during the period<sup>29–31</sup>. Although SE constitutes the majority of the endosperm by volume, SE is rather simple in its structure compared to other cell types. From 8 DAP to 16 DAP stages, SE cells mainly undergo cell proliferation<sup>7</sup>.

We employed several criteria to identify noncoding transcripts. First, possible partial sequence reads and artificial sequencing reads were removed based on FPKM values. Only the transcripts with an FPKM value more than 0.1 were retained. We then selected transcripts that had not been annotated in the B73 maize genome assembly<sup>13</sup>. The transcripts that overlapped with any known protein coding transcript were also excluded, as they may represent alternatively spliced variants. Remaining transcripts were filtered based on the coding potential using the transdecoder program<sup>32</sup>. A total number of 5,793 transcripts passed the criteria and were categorized as noncoding transcripts (Table S2). Similar to the distribution of the coding transcripts, noncoding transcripts were more abundant in BETL than in the other two cell types (Fig. S1B), suggesting that BETL displays wider transcriptional diversity than the other two cell types of the endosperm.

Among various noncoding transcripts, our emphasis was to catalogue novel long noncoding transcripts that may play regulatory roles in the endosperm cell differentiation. For this purpose, we applied two more filters to shortlist novel long noncoding transcripts. First, we removed transcripts shorter than 200 nucleotides. Second, we removed transcripts predicted to be small RNA precursors (such as miRNA and siRNA) or transposable elements. A total of 1,540 transcripts that satisfied all criteria were chosen for further analysis (Figs 1D and S1B).

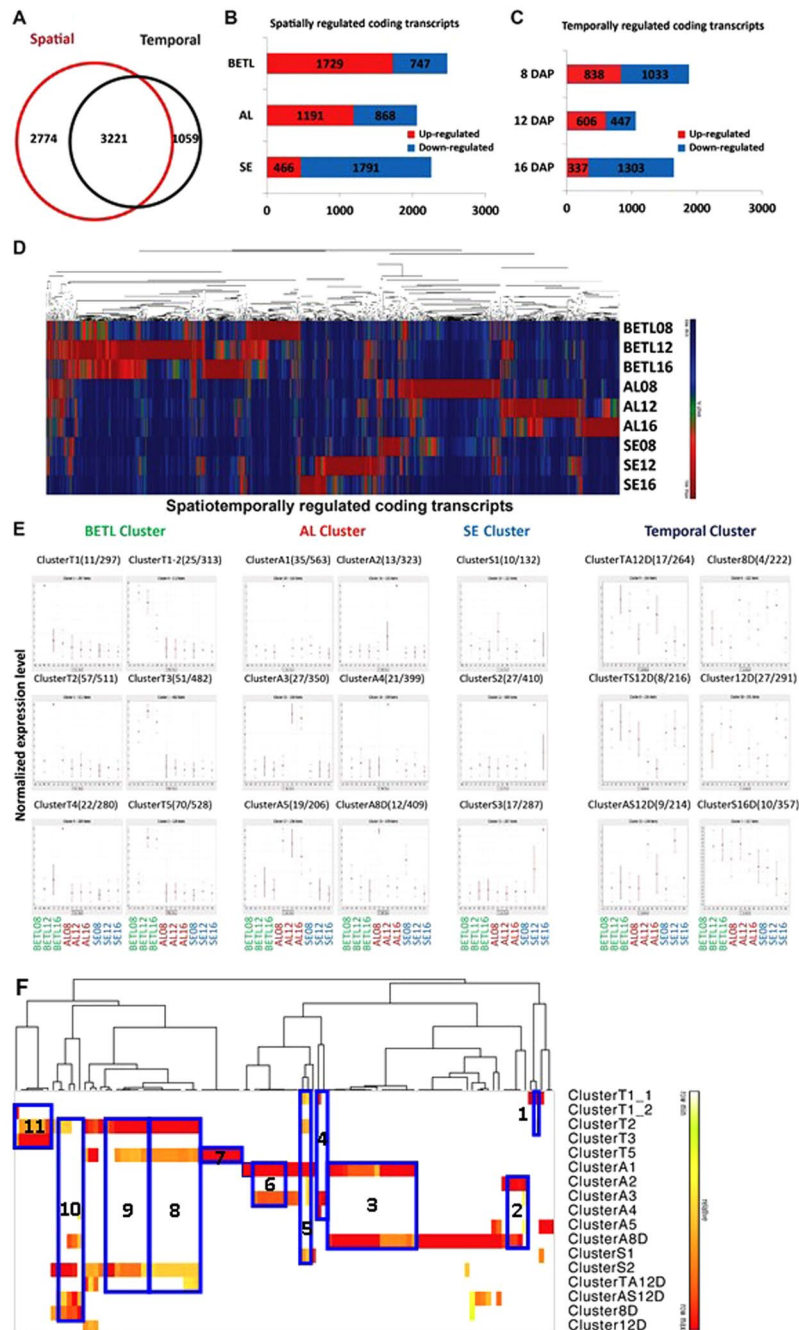
**Identification of spatio-temporally regulated protein-coding transcripts.** To characterize cell-type-specific patterns of expression of coding transcripts, we identified protein-coding transcripts that were highly enriched or depleted in a particular cell type. Transcripts that were up-regulated or down-regulated more than 4-fold in a single cell type are termed “spatially-regulated coding transcripts”. Similarly, coding transcripts that were either enriched or reduced at a particular developmental stage by more than 4-fold are named “temporally-regulated coding transcripts”. We identified a total number of 7,054 coding transcripts that are spatio-temporally regulated; 5,995 coding transcripts were spatially regulated and 4280 were temporally regulated (Fig. 2A). Among the spatio-temporally regulated coding transcripts, ~46% (3,221) were localized to one of the three tissue types and exhibited stage-specific expression patterns (Fig. 2A). Approximately 39% (2,774) of spatio-temporally regulated transcripts were confined to a single tissue type without stage-specific expression. Further, 1,059 (~15%) of spatio-temporally-regulated transcripts were up-regulated in a particular developmental stages but were detected in all endosperm tissue types (Fig. 2A).

Among the tissue-specific coding transcripts, 3,386 transcripts were up-regulated at one of the three-time points, and over 50% of those up-regulated genes were confined to BETL (Fig. 2B). AL and SE contained 35% and 15% of the up-regulated coding transcripts, respectively. Up-regulation patterns of coding transcripts exhibited significant shifts at each developmental stage (Fig. 2C). Transcription of protein-coding genes was most active in BETL at 8 DAP (Fig. 2D). The large number of tissue-specific coding transcripts and their oscillating patterns during maize endosperm development indicate that a transcriptional regulatory network plays a crucial role in functional differentiation of cells in the three tissues.

**Characterization of spatio-temporally regulated protein-coding transcriptome clusters.** We further classified coding transcripts that exhibit spatio-temporal regulation by clustering transcripts according to spatio-temporal expression patterns to identify co-regulatory modules. Because spatio-temporal transcriptome profiles are complex and multi-dimensional, we utilized the self-organizing map (SOM) analysis tool to reveal underlying patterns. Twenty-one spatio-temporally specific clusters were uncovered by the SOM analysis (Fig. 2E). Among the 21 clusters, 15 clusters display obvious tissue type-specific patterns, including six BETL up-regulated groups, six AL up-regulated groups, and three SE up-regulated groups. Among these tissue-specific groups, ten clusters include transcripts concentrated at a single spatio-temporal point (Fig. 2E), indicating that these groups of tissue-specific genes are under tight regulation during endosperm cell differentiation.

To assign biological functions to clustered coding transcripts within distinct cell types and developmental stages, we examined Gene Ontology (GO) term enrichment using the hypergeometric distribution test and subsequent correction by the multiple hypothesis test<sup>33</sup>. Enriched coding transcripts with *q*-values less than 0.05 were used for the GO term enrichment analysis, and expression values of transcripts were converted into the log<sub>10</sub> scale. Significantly-enriched functional categories of each cluster were tabulated and summarized by the hierarchical cluster analysis (Fig. 2F).

Functional GO-term enrichments of 17 clusters out of 21 clusters showed distinct trends. First, distinct biological functional groups are enriched in cell-type specific clusters, especially BETL and AL clusters. In BETL clusters, defence responses, proteolysis, post-translational related processes, transport, and localization groups were far more abundant than in the other two cell types throughout the three developmental stages (Fig. 2F). The lipid metabolic process and proteolysis were abundant functional groups in BETL up-regulated clusters at 8 DAP. Defense responses and stress-related groups were observed in the 8 DAP up-regulated cluster, whereas post-transcriptional regulation-related functional groups, such as protein modification and phosphorylation, are enriched in up-regulated clusters of later developmental stages (12 DAP and 16 DAP). Across all developmental stages, coding transcripts involved in intracellular transport and localization are concentrated in clusters of genes up-regulated in the BETL relative to other cell types. It has been shown that BETL accumulates anti-pathogen



**Figure 2.** Differential expression of coding transcripts during endosperm development. **(A)** Coding transcripts with FPKM > 1.0 and with more than 4-fold expression-level changes were analysed; the overlap between spatially and temporally regulated are shown. **(B)** The numbers of coding transcripts that were up- or down-regulated more than 4 fold in any tissue are shown in box graphs. **(C)** The numbers of coding transcripts that were up- or down-regulated more than 4 fold at any DAP are shown in box graphs. **(D)** Hierarchical cluster analysis of spatio-temporally regulated coding transcripts. **(E)** SOM clusters are depicted based on clustering patterns. Given in parentheses is the number of co-expressed transcription factors and the number of genes that belong to each SOM cluster. The relative level of transcript is plotted on the y axis for each of nine samples: BETL08, BETL12, BETL16, AL08, AL12, AL16, SE08, SE12, and SE16 from left to right. **(F)** Functional GO-term enrichment. Significance is represented by heat map ( $-\log_{10}$  P-value; red – high/white – low). Representative functional categories correspond to numbers on graph as follows: 1, defence response; 2, oxaocid metabolic process (oxidation reduction); 3, macromolecular complex subunit organization (chromatin assembly or disassembly, nucleosome assembly, protein-DNA packing); 4, lipid biosynthesis processes; 5, response to stimulus and stress; 6, response to abiotic stimulus and biotic stress (temperature); 7, microtubule-based process; 8, post translational process (phosphorylation, protein modifications); 9, regulation of metabolic/ biosynthesis process; 10, regulation of transcription process; 10, macromolecule biosynthesis process; 11, establishment of localization.

proteins and secretes small soluble proteins to protect maternal tissues from pathogens<sup>7, 11, 12, 34</sup>, thus, in this regard, BETL GO-term clusters are in agreement with previous studies.

The AL is the outermost layer of the endosperm<sup>7, 35</sup>. The enrichment for transcripts associated with microtubule-dependent movements and responses to abiotic stresses (Fig. 2F) is consistent with the fact that the AL differentiation involves active cell division and cell enlargement to accommodate the enlarging endosperm during development. Functional clusters of lipid biosynthesis processes are first observed in AL at 8 DAP, and their occurrences increase at 12 and 16 DAP. The oxidation-reduction and carbohydrate metabolic processes, which are related to lipid biosynthesis, were observed in AL up-regulated clusters at all developmental stages. These transcriptome signatures confirm the function of AL in which lipid bodies accumulate during seed development. In addition, AL up-regulated clusters at 8 DAP include transcription regulation via chromatin alteration and protein complex assembly (Fig. 2F), suggesting developmental reprogramming occurs at early AL differentiation.

The main function of SE is to store starch and seed proteins<sup>7, 35</sup>. As expected, cellular carbohydrate metabolic process, responses to abiotic stimuli and cellular polysaccharide metabolic processes were most significantly enriched GO terms in SE up-regulated coding transcript clusters.

As discussed above, the majority of distinctive GO-term functional groups in each cluster were observed among cell-type-specific SOM clusters. Some GO-term functional groups included transcripts expressed across multiple cell types, however. Enriched functional groups associated with transcriptome clusters in multiple cell types are involved in more general biological processes. For example, genes involved in cellular macromolecule metabolic processes are up-regulated in both SE- and AL-specific clusters, and enrichment in GO functional groups involved in regulation of gene expression and metabolic process was observed in both BETL- and SE-specific clusters. Carbohydrate metabolic process functional group was enriched in all tissue-specific clusters. The enriched functional groups in SE frequently overlapped with one or two tissue specific clusters. These enriched functional groups represent general metabolic and biosynthetic pathways. Taken together, distinct GO-term enrichments of spatio-temporally regulated genes indicate that coding transcripts from compartmentalized endosperm cells participate in different biological pathways that contribute to endosperm development in at distinct times and in certain regions of the organ.

### Spatio-temporally regulated transcription factor expression in developing maize endosperm.

Transcription factors (TFs) play critical roles in specifying gene expression during development. Therefore, we extracted the list of TFs from the spatio-temporally regulated clusters (Figs 2E and S2). Of 2,605 annotated TFs, 492 were represented in the SOM clusters that show spatio-temporal expression patterns (Fig. 2E). All 21 clusters included sets of co-expressed TFs, indicating that various TFs show distinct expression patterns over the course of endosperm development. Of the spatio-temporally clustered TFs, 42% belong to BETL clusters, whereas AL and SE clusters account for 31% and 27%, respectively (Fig. S2A). Among 492 spatio-temporally regulated TFs, 72% were up-regulated and 28% were down-regulated in a spatio-temporal manner (Fig. S2A,B). In particular, in BETL at 8 DAP, the transcriptome includes the majority of spatio-temporal-specific TFs, suggesting that dynamic transcriptional reprogramming occurs in BETL at around this time point after pollination (Fig. S2C–E). We validated the expression patterns of selected transcription factors by qRT-PCR (Fig. S2F).

**Spatio-temporally regulated imprinted genes.** Genomic imprinting is involved in plant endosperm development. Using data from several studies on parent-of-origin specific gene expression profiles in the maize endosperm<sup>22–24, 26</sup>, we collated a list of imprinted protein-coding genes. Subsequently, we examined whether expression of genes on this list was spatio-temporally regulated during endosperm development. Our coding transcriptome data set includes 67% (247/370) of known maternally expressed imprinted genes (MEGs) and 61% (184/298) of known paternally expressed imprinted genes (PEGs). Among sequenced imprinted genes, 75% of MEGs and 32% of PEGs belong to spatio-temporally regulated transcript clusters (Fig. S3A,B). Interestingly, both MEGs and PEGs with spatio-temporally regulated patterns were over-represented in the BETL transcriptome at 8 DAP (Fig. S3C,F). Spatio-temporally regulated MEGs also enriched in the AL transcriptome at 12 DAP (Fig. S3C,F). The majority of cell-type-specific PEGs (30/54) are expressed at the highest level in BETL, although there is no clear bias to a certain developmental stage (Figs 5F, S3D and S4B).

### Spatio-temporally regulated H3K27me3 target loci in developing maize endosperm.

Chromatin-state signatures are often associated with transcriptional activity. Histone H3 trimethylation at lysine 27 (H3K27me3) is one cell-type-specific epigenetic marker in higher eukaryotes<sup>25, 36, 37</sup>. We examined whether H3K27me3-enriched regions corresponded to spatio-temporally regulated coding transcripts by chromatin immunoprecipitation (ChIP) followed by sequencing (Fig. S5). About 18% (306/1756) of H3K27me3-enriched regions overlapped with spatio-temporally regulated coding transcript clusters (Fig. S5A–D). Transcripts from H3K27me3-enriched loci were over-represented in BETL compared to other tissues, as BETL clusters included 45% of spatio-temporally regulated H3K27me3-marked loci (Fig. S5A). Among these temporally regulated transcripts, 44% of them were expressed at 8 DAP (Fig. S5B), suggesting the active chromatin remodelling occurs at the early developmental stage. The expression of transcripts from some of the H3K27me3-enriched loci was validated by qRT-PCR (Fig. S6).

**Identification of spatio-temporally regulated long noncoding RNA (lncRNA).** As described earlier, our sequencing reads allowed us to identify a large collection of novel long noncoding RNAs (lncRNAs). lncRNAs are evenly distributed in the maize genome with the exception of slightly lower lncRNA density in chromosome 1 compared to the other chromosomes (Fig. 3A). This pattern is similar to the distribution of coding transcripts. The majority of lncRNAs are shorter in length than coding RNAs, with few (4.2%) longer than

1.1 kb. In comparison, the majority (89%) of coding transcripts were longer than 1.1 kb (Fig. 3B). To identify spatio-temporally regulated lncRNAs, we employed the same criteria used to identify spatio-temporally regulated coding transcripts (Fig. 2). Interestingly, more than 85% (1,312) of sequenced lncRNAs showed spatio-temporal expression patterns (Fig. 3C). It is noteworthy that most spatio-temporally regulated lncRNAs (over 90%) were expressed only in a certain tissue or at a certain developmental stage (Fig. 3D,E).

Among the lncRNAs, 82% were to spatio-temporally regulated, and the rest of the lncRNAs are either tissue-specific (11%) or developmental-stage specific (7%) (Fig. 3C,F). To better understand spatio-temporal patterns, an SOM analysis was used (Fig. 3G). Fourteen spatio-temporally regulated clusters were identified, with six BETL-specific clusters, four AL-specific clusters, and four SE-specific clusters. Among the spatially regulated lncRNAs, 51%, 21%, and 27% of lncRNAs were BETL-, AL-, and SE-specific lncRNAs, respectively. The spatio-temporal-specific regulation of lncRNA occurs mostly in BETL. Among temporally regulated lncRNAs, 51%, 28%, and 21% of lncRNAs were specifically expressed at 12 DAP, 8 DAP, and 16 DAP, respectively. Therefore, the BETL transcriptome at 12 DAP had the most spatio-temporally-regulated lncRNAs and coding RNAs of the endosperm tissues examined. Unlike the coding transcript cluster, however, 98% of spatio-temporally regulated lncRNAs were up-regulated only at in a certain tissue and/or at a certain developmental stage (Fig. 3D–G), indicating that the expression of lncRNAs is highly specific.

**Spatio-temporally regulation of H3K27me3 at loci encoding lncRNA.** Many lncRNAs have been implicated in mediating chromatin modification in eukaryotes<sup>38–40</sup>. To identify the lncRNAs that correlate with chromatin modification in endosperm cells, we compared the relative distributions of H3K27me3-enriched regions and lncRNA loci. About 30% of spatio-temporally regulated lncRNA loci overlapped with H3K27me3-enriched loci (Fig. 4A). The Chi-square test confirmed that cell-type-specific lncRNAs and developmental-stage-specific lncRNAs significantly overlapped with H3K27me3-enriched loci (Fig. S7). The majority of H3K27me3-enriched lncRNA loci were over-represented in BETL compared to the other tissues (Fig. 4B). Among developmental stage-specific H3K27me3-enriched lncRNAs, 45% of them showed 12 DAP specificity (Fig. 4C). Significant overlaps between lncRNAs and H3K27me3-enriched loci suggest that certain lncRNAs contribute to the enrichment of H3K27me3 and/or that the regulation of lncRNAs themselves are under the control of H3K27me3 (Fig. 4D). We validated the ChIP results from certain lncRNA loci by RNA *in situ* hybridization and qRT-PCR (Fig. S8).

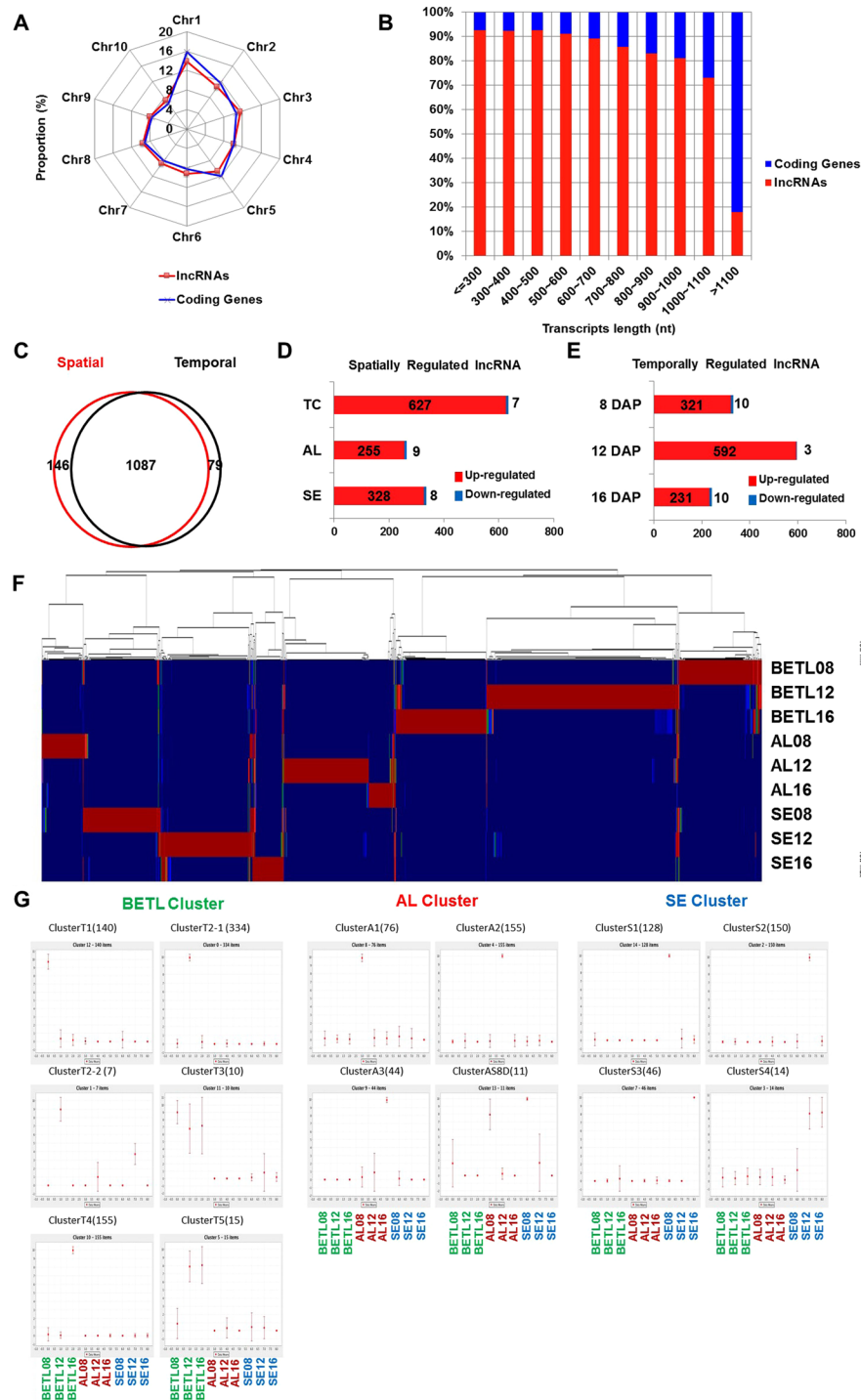
**Spatio-temporal expression patterns of certain lncRNAs and neighbouring coding transcripts are correlated.** Transcription of coding genes is often regulated by neighbouring lncRNAs, and the transcription of lncRNAs can be influenced by the transcription of neighbouring coding genes<sup>41</sup>. To address whether the spatio-temporal expression pattern of lncRNAs is correlated with the expression of neighbouring coding genes, we compared expression patterns of each lncRNA with coding genes within 10 kb up- and downstream of lncRNA loci. In the majority of cases, there was no clear correlation between the expression of coding and noncoding transcripts (Fig. 5A), suggesting that the majority of lncRNAs loci are under their own transcriptional control.

We did observe overlap between coding and non-coding transcripts in 11 spatio-temporal expression clusters (Figs 2E and 3G). Therefore, we examined the presence of shared *cis* elements between coding and noncoding transcripts in these common spatio-temporal transcriptome clusters (Fig. 5B). We examined 5- to 22-bp sequence elements using the MEME software<sup>42</sup> and the TOMTOM program<sup>43</sup> to identify shared *cis*-regulatory modules with significant similarity ( $q$ -value < 0.05) between coding and noncoding transcripts of common spatio-temporal clusters (Fig. 5C). Nine sets of clusters exhibited significantly shared *cis* elements (Table S3). Most over-represented *cis*-regulatory motifs were found in BETL clusters of coding (38%) and noncoding (50%) transcripts. Clusters shared from 23% to 94% of total *cis*-regulatory modules. With the exception of the A3 cluster, all tissue-specific clusters of lncRNA included shared *cis*-regulatory motifs with coding genes that were clustered in similar SOM patterns. AP2/ERF binding motifs were found in all shared clusters of both coding and noncoding clusters, similar to a result using only coding transcriptome<sup>44</sup>. Therefore, the AP2/ERF transcription factor family may play a broad role in the transcriptional regulation of both coding and long noncoding transcripts during maize endosperm development.

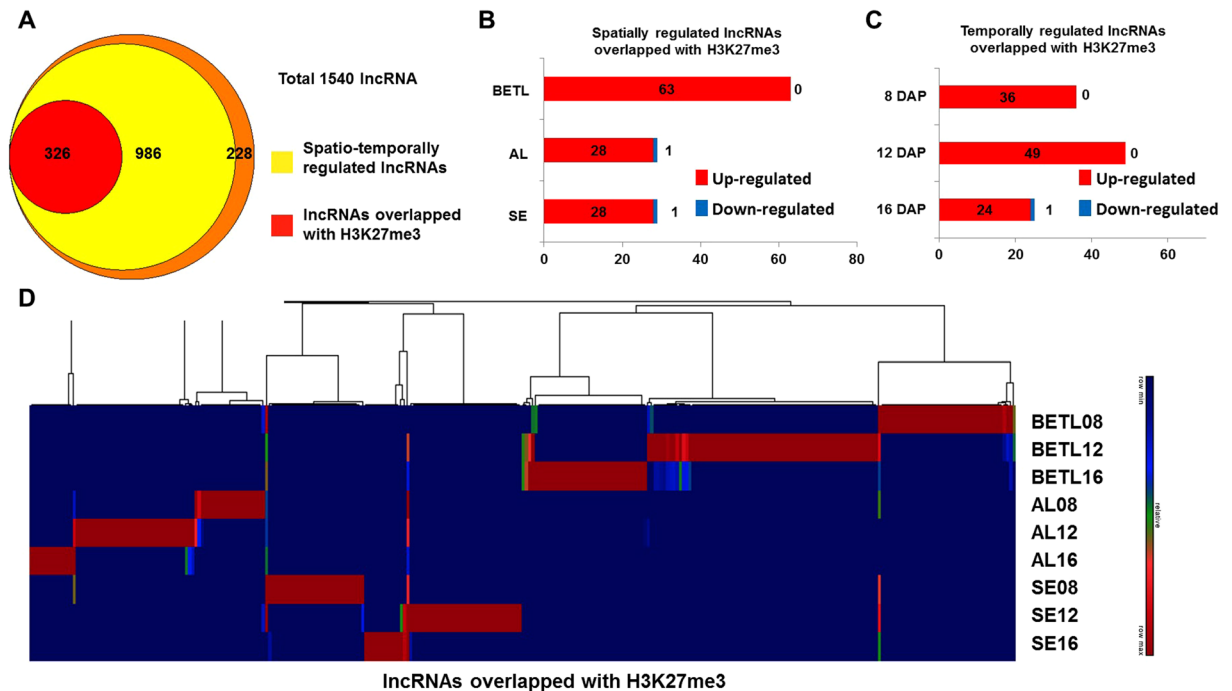
**Histone modifications influence expression of spatio-temporally regulated loci.** To verify the involvement of histone modifications in the differential expression of spatio-temporally regulated transcripts, we examined several BETL-specific coding and long noncoding loci for mRNA expression, histone H3 lysine 4 trimethylation (H3K4me3) enrichment, and H3K27me3 enrichment. The differential enrichment of H3K4me3 and H3K27me3 during development has been shown in a number of developmentally regulated genes in eukaryotes, and these modifications are mediated by the Trithorax-Polycomb regulatory networks. H3K4me3 markers correlate with gene activation, whereas H3K27me3 markers correlate with gene repression<sup>45</sup>. As expected, we observed that the enrichment of H3K4me3 increased when transcripts were up-regulated, whereas the enrichment of H3K27me3 increased when transcripts were down-regulated (Fig. 6). Interestingly, lnc2653 also exhibited similar correlative patterns of histone modifications, suggesting that the regulation of lncRNA expression is similar to that of protein-coding transcripts and that the Trithorax-Polycomb complex mediates the spatio-temporal regulation of gene expression during endosperm development.

## Discussion

In this study, we comprehensively profiled the transcriptomes of the maize endosperm at three developmental time points in three different tissues. Previous transcriptome analyses of the maize endosperm were performed



**Figure 3.** Identification of long noncoding RNAs in the maize endosperm. **(A)** Distribution of lncRNAs along each chromosome. The proportion of coding transcripts and lncRNAs for each chromosome. **(B)** Size distributions of lncRNAs and coding transcripts. **(C)** Venn diagram of the number of spatially and/or temporally regulated lncRNAs with FPKM >0.01 and more than 4-fold expression-level changes. **(D)** The numbers of lncRNAs up- or down-regulated more than 4-fold in any tissue are shown. **(E)** The numbers of lncRNAs up- or down-regulated more than 4 fold at any DAP are shown. **(F)** Hierarchical cluster analysis of spatio-temporally regulated lncRNAs. **(G)** SOM clusters depicted based on clustering patterns. In parentheses are the number of co-expressed transcription factors and the number of genes that belong to each SOM cluster. The relative level of transcript is plotted on the y axis for each of nine samples: BETL08, BETL12, BETL16, AL08, AL12, AL16, SE08, SE12, and SE16 from left to right.



**Figure 4.** Spatio-temporally regulated lncRNA loci overlap with H3K27me3-marked loci. **(A)** Venn diagram representing the overlap between H3K27me3-enriched loci and spatio-temporally regulated lncRNA transcription sites. **(B)** The numbers of lncRNAs that overlap with H3K27me3 enriched loci that are up- or down-regulated more than 4-fold at any tissue are shown. **(C)** The numbers of lncRNAs that overlap with H3K27me3-enriched loci that were up- or down-regulated more than 4 fold at any DAP are shown. **(D)** Hierarchical cluster analysis of spatio-temporally regulated lncRNAs that overlap with H3K27me3-enriched loci.

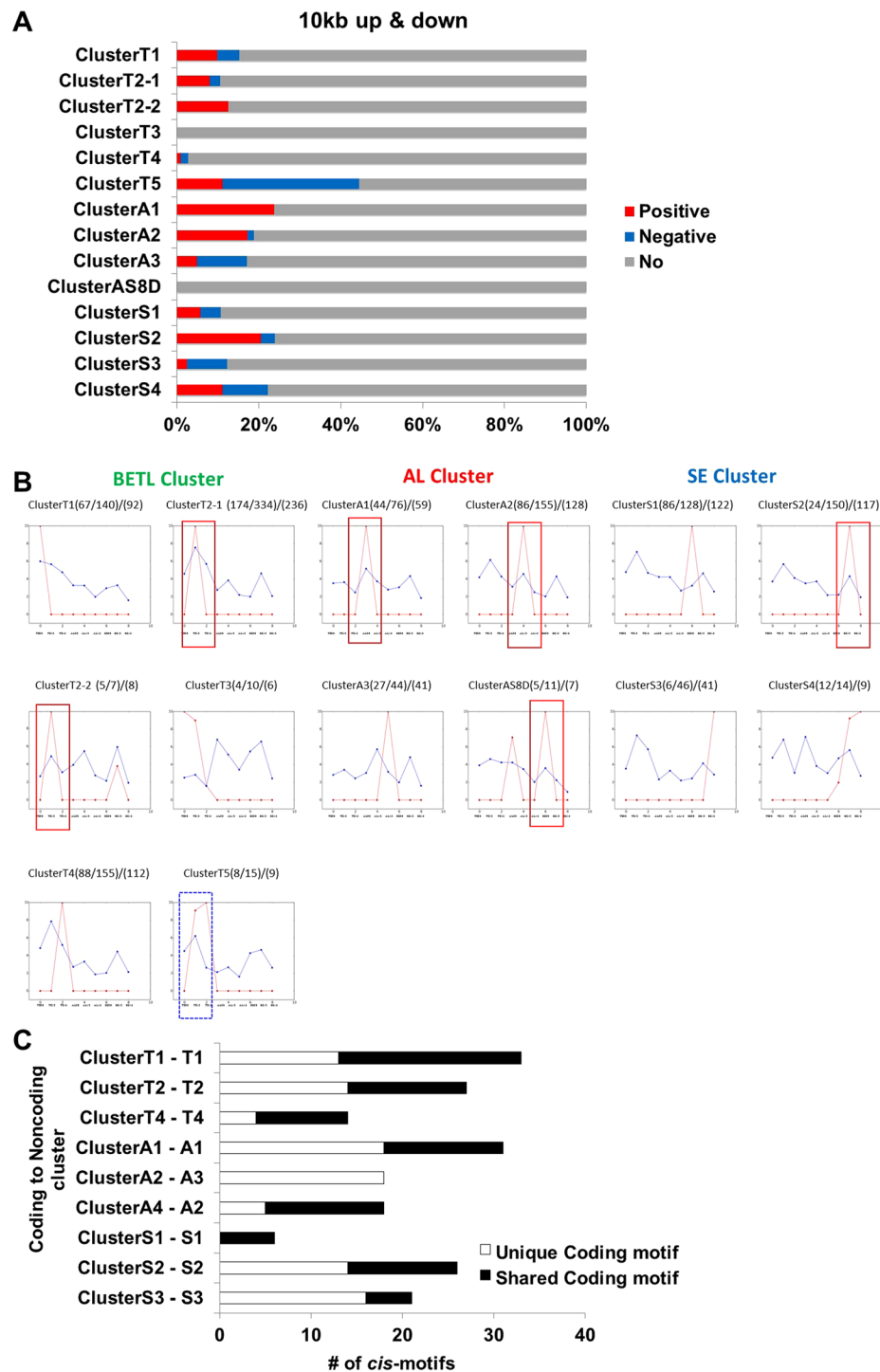
on transcripts obtained from dissected endosperm tissues at a certain developmental stage<sup>10, 16, 18, 19, 22, 23, 25, 26</sup> or from whole endosperms from different developmental stages<sup>17, 20, 21, 24</sup>. The maize endosperm is suited for a combinatorial cell-type-specific and developmental analysis because of its compartmentalized architecture and highly differentiated nature of its cellular constituents<sup>46</sup>. Our study is the first to reveal the two-dimensional transcriptomic landscape of maize endosperm development.

Our analysis revealed 7,054 spatio-temporally regulated protein-coding transcripts, indicating that massive transcriptional regulations are involved in maize endosperm development. In addition, we identified 1,540 novel long noncoding transcripts and found that 1,312 were spatio-temporally regulated, suggesting that lncRNAs make important contributions to development. Using SOM analysis we classified these spatio-temporally regulated transcripts into co-expressed clusters and demonstrated that co-expressed coding and noncoding transcripts mirror the stage of endosperm differentiation within each cell type. In particular, the co-expressed clusters of coding transcripts exhibited distinct functional category enrichment patterns that are consistent with the developmental stages and the compartmentalized tissues in the endosperm. A total of 492 TFs were found in spatio-temporal clusters. These TFs may control other transcripts in the same cluster. For example, MRP-1 controls a subgroup of BETL-specific genes through a GATA-containing motif<sup>44</sup>. Spatio-temporally regulated TFs identified in our study are candidates for key transcriptional regulators of the endosperm development.

The abundance of genes displaying a high degree of transcriptional specificity at 8 days after pollination indicates that cell fates have been determined before 8 DAP (2). The genes that we have identified are related to maturation and maintenance of the differentiated cells as opposed to cell fate determination. Transcriptome profiling of endosperms at earlier stages is required to investigate expression of genes that determine cell fate.

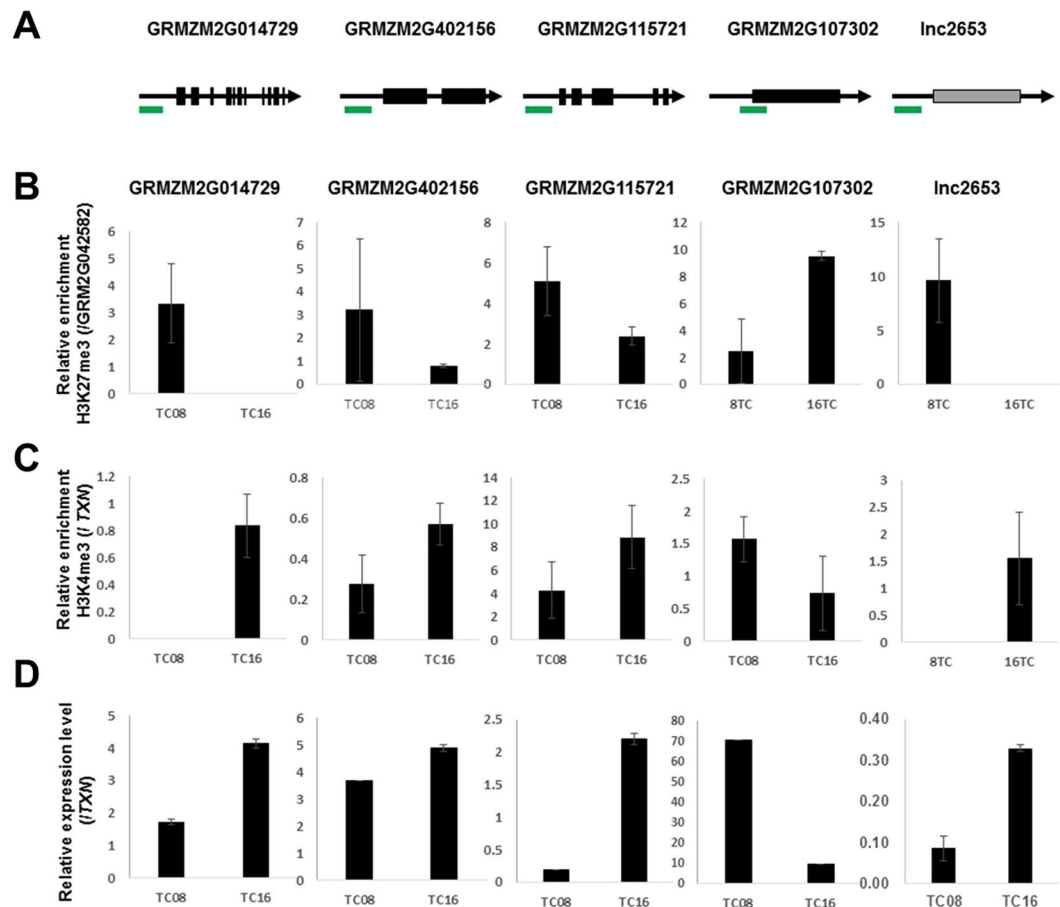
In plants, imprinting is prevalent in the endosperm<sup>47</sup>. Previous studies have investigated imprinting in the endosperm or embryo as a whole. It is possible that mono-allelic expression can lead to the activation or repression of one allele in a cell-type-specific manner. Although our study could not examine allele-specific expression, we provide evidence that maternally and paternally expressed imprinted genes are not uniformly distributed over the endosperm. MEGs and PEGs identified from 12 DAP whole endosperm samples are available<sup>19, 22, 23, 26</sup>, and by comparison of our data with this previously collected data, we found that both MEGs and PEGs are most abundant in BETL clusters (Figs S3 and S4). This suggests that monoallelic regulation of either MEGs or PEGs is concentrated in the BETL. The spatio-temporal bias is more obvious in MEGs (Fig. S4), with the majority of MEG expression observed at 12 DAP in BETL (Fig. S3). The parental conflict hypothesis is a generally accepted explanation for genomic imprinting<sup>48</sup>. The hypothesis predicts that genes regulating growth of progeny exhibit parent-of-origin expression. *Meg1* is an example of maternally expressed imprinted gene expressed in BETL and involved in seed development<sup>49</sup>. Given that the rate of nutrient allocation to developing seeds is determined by





**Figure 5.** Co-regulatory relationships between noncoding and coding transcripts. **(A)** Correlations between expression of lncRNA and neighbouring coding genes within 10 kb upstream and downstream of lncRNAs transcript loci were calculated for lncRNAs in each cluster. **(B)** The median value of an expression of lncRNAs and their neighbouring coding genes in each cluster. The relative level of transcript is plotted on the y axis for each of nine samples: BETL08, BETL12, BETL16, AL08, AL12, AL16, SE08, SE12, and SE16 from left to right. Blue lines plot data for coding genes and red lines lncRNAs. Red boxes indicate positive correlations and blue box indicates negative correlation. **(C)** The number of enriched cis-regulatory modules between each SOM cluster of lncRNAs and coding genes showing similar expression patterns.

maternal to filial transport through BETL, the abundance of imprinted genes in BETL agrees with the parental conflict hypothesis.



**Figure 6.** Correlation among RNA expression, H3K27me3, and H3K4me3 in BETL at 8 DAP (TC08) and 16 DAP (TC16). (A) Schematic representation of the locations of amplicons used for ChIP assays. (B) Spatio-temporal enrichment of H3K27me3 at indicated loci plotted relative to GRMZM2G042582. (C) Spatio-temporal enrichment of H3K4me3 at indicated loci plotted relative to *TXN* (GRMZM2G066612). (D) mRNA expression at indicated loci relative to expression of *TXN* (GRMZM2G066612).

Recent studies have shown that lncRNAs play various roles in the regulation of gene expression during development in eukaryotes, from chromatin remodelling to transcriptional interference. Several lncRNAs have been shown to function in the regulation of gene expression in plants<sup>39,40,50</sup>. Other previous works examined various organ-specific transcriptomes and epigenomes<sup>51,52</sup>. In our analysis, we isolated tissues by cryo-dissection and analysed lncRNA expression. Spatio-temporal expression patterns were detected for 85% of identified lncRNAs, and over 90% of lncRNAs are expressed only in a certain tissue or at a certain developmental stage (Fig. 5), supporting their roles in spatio-temporal regulation during endosperm development. Our data show that expression patterns of most lncRNAs are not correlated with those of their neighbouring genes (Fig. 5A), suggesting that lncRNA transcription is not induced by neighbouring transcriptional activity. We did, however, identify *cis* motifs that are shared among similarly clustered coding and noncoding loci (Fig. 5B,C). Therefore, lncRNAs expression appears to be transcriptionally regulated but independently of neighbouring genes. Biased expression of lncRNAs during endosperm development implies that these RNAs perform a necessary function. Interestingly, 30% of spatio-temporally regulated lncRNA loci overlap with H3K27me3-enriched loci (Fig. 4). H3K27 trimethylation is mediated by the Polycomb Repressive Complex 2 (PRC2), and our data suggest that some lncRNAs might contribute to the enrichment of H3K27me3 and/or the regulation of lncRNA itself under the control of H3K27me3. For example, we observed that the repression of lnc2653 correlates with the increased enrichment of H3K27me3 (Fig. 6).

Identification of cell-type-specific lncRNAs in our study provides evidence that lncRNAs are critical factors in differentiation of the maize endosperm. Considering that lncRNAs control chromatin structures, epigenetic mechanisms are likely to play roles in cell differentiation. It is possible to isolate chromatin specimens for ChIP experiments from the cryo-dissected endosperm tissues, as demonstrated in Fig. 6. A large-scale ChIP-Seq analysis of BETL, AL, and SE in parallel with profiling lncRNA populations is a possible direction for characterizing chromatin modification associated with cell differentiation in the plant endosperm.

## Materials and Methods

**Sample preparation and RNA analysis.** All kernel samples were harvested from self-pollinated B73 ovules. Kernels were collected at 8, 12, 16 days after pollination (DAP). All collected samples were frozen immediately in liquid nitrogen and stored at  $-80^{\circ}\text{C}$ . Kernels were embedded in Tissue-Tek optimum cutting temperature-embedding compound, mounted on sectioning stub, and cut into 15–30 micrometer sections in a cryostat precooled to  $-22^{\circ}\text{C}$  (Leica Microsystems). The cryo-sectioned samples were dehydrated and stained with a HistoGene LCM Frozen Section Staining Kit according to the provided manufacturer's protocol. Approximately 150 sections of each tissue were dissected manually under a microscope, followed by RNA extraction<sup>12</sup>. Total RNA was isolated using Picopure RNA isolation Kit (Arcturus) according to the manufacturer's instructions. Genomic DNA was removed by DNase (Qiagen) treatment performed according to instructions supplied with the kit. The integrity of extracted total RNA samples was tested analysis by polyacrylamide gel electrophoresis and using an Agilent 2100 Bioanalyzer. Possible contaminations were also examined by quantitative real-time polymerase chain reaction (qPCR) with standard cell-type-specific markers<sup>12</sup>. The RNA was further purified over a Qiagen RNeasy column. The extracted total RNA was used for relative transcript quantification by quantitative RT-PCR and high-throughput RNA-Seq analyses.

**Library preparation and sequencing.** To construct RNA-Seq libraries, we first depleted samples of ribosomal RNA using the RiboMinus Plant Kit (Thermo Fisher). This is necessary to achieve optimal coverage, good detection sensitivity, and reliable results. PolyA-enriched RNA was then captured using oligo(dT) beads (Invitrogen) and was fragmented. The first-strand cDNA was synthesized by reverse transcription of the fragmented PolyA-enriched RNA, followed by second strand synthesis with DNA polymerase to generate double-stranded DNA. After end-repair and A-tailing, barcode adapters were ligated. After PCR amplification, DNA was captured on AMPure beads to purify and remove adaptor dimers. Each barcoded RNA-Seq library was sequenced on an Illumina HiSeq2500 by the University of Texas at Austin Core facility. An average of 20 million reads of  $2 \times 200$  nucleotides paired-end sequences were generated. Prior to mapping, the sequencing reads were examined by FastQC software to remove the low-quality reads and adaptors. The remaining reads were aligned to the maize genome (ZmB73\_RefGen\_v2<sup>13</sup> using Tophat<sup>27</sup>. More than ~80% of the sequencing reads from each of the nine libraries were mapped to the reference genome (Table S1). The aligned reads were then subjected to the Cufflinks assembly<sup>28</sup> using ZmB73\_5b\_filtered set and FPKM values were calculated. The coding transcripts with FPKM value  $> 1$  and the noncoding transcripts with FPKM value  $> 0.1$  were used in further analyses.

**lncRNA identification.** Read alignment to the reference genome was performed using the Bowtie, Tophat, and Cufflinks pipeline. The transcripts matched to annotated transcripts were filtered out using SAMtools and a customized awk script. The novel transcripts from intergenic regions were further analysed. The individual assembled novel transcripts from each library were merged into one non-redundant set using Cuffmerge option. The merged transcript assembly (final GTF file) was used in quantification of expression levels using Cufflink. Putative lncRNAs were arbitrarily defined as transcripts shorter or longer than 200 nt with lack or weak protein coding ability. The novel transcripts shorter than 200 nt were excluded using customized scripts. More than 95% of protein coding genes have open reading frames (ORFs) of more than 100 amino acids<sup>33</sup>. The novel long transcripts with long ORFs were removed using a coding potential program (TransDecoder tool) and customized script. To remove putative precursors of all small RNA<sup>54</sup> and transposable element (TE) transcripts (<http://maizetdb.org/~maize/>), putative lncRNAs were aligned to small RNA and TE databases and those identical to small RNAs and TEs were removed using Python<sup>54</sup> and customized scripts.

**SOM analysis.** The normalized FPKM values of spatio-temporally regulated transcripts were used for self-organizing maps (SOM) analysis. SOM clustering was performed with SOM module embedded in the GenePattern based on the gene expression level (<http://www.broadinstitute.org/cancer/software/genepattern>). Clustering analysis was performed five times with 4,200,000 repetitions of the process, subsequently generating 21 clusters for coding transcripts and 14 clusters for lncRNAs.

**Identification of *cis*-regulatory elements.** Genomic sequences were obtained from 1 kb upstream of transcription start sites of coding transcripts. For lncRNAs, the genomic sequences were extracted from 1 kb upstream of the transcription start site and 1 kb downstream of the transcription end site using customized awk scripts and Bedtools. The MEME program was used to identify *de novo* motifs in those regions. Motifs with E-value lower than 0.01 were analysed using the TOMTOM motif comparison tool. Only those with q-values lower than 0.01 were taken for further analysis to identify motifs with significant similarity between coding RNA and lncRNA clusters.

**Chromatin immunoprecipitation and ChIP-qPCR.** The cryo-sectioned tissue samples were directly cross-linked in 1% formaldehyde-containing buffer. Crosslinked tissue samples were ground in liquid nitrogen and used in ChIP experiments as previously described<sup>55</sup>. Antibodies recognizing H3K4me3 (Abcam cat. no.: ab8580) and H3K27me3 (Abcam cat. no.: ab6002) were used for ChIP assays. Enrichment of each fragment was first calculated based on comparison to quantitative PCR (qPCR) of input material. Second, relative enrichment was normalized by comparison to each listed reference gene (*TXN* for H3K4me3 and *GRMZM2G042582* for H3K27me3). Lastly, to show relative fold change, values of reference genes were converted to 1; relative fold changes are presented in figures. Information on primer sequences used in ChIP-qPCR analyses are shown in Supplemental Table 3. qPCR was performed using Maxima SYBR green master mix (Thermo Scientific) according to the manufacturer's instructions. All ChIP-qPCR analyses were performed with at least two biological

replicates and with two or three technical replicates. qPCR was performed in a 384-well PCR plate using the ViiA™ 7 Real-Time PCR System (Life Technologies).

**Quantitative RT-PCR.** First-strand cDNA synthesis was performed with 1 µg of total RNA and oligo(dT) for coding genes or random primers for noncoding RNAs using M-MLV Reverse Transcriptase (Promega). Real-time qRT-PCR was carried out using Maxima SYBR Green/ROX qPCR Master Mix (Thermo Scientific) on a ViiA7 Real-Time PCR system (Applied Biosystems). The PCR cycling conditions were as follows: 95 °C for 10 min, 40 cycles of 95 °C for 15 sec, 60 °C for 30 sec, and 72 °C for 30 sec. The relative transcript levels of coding genes and lncRNAs were determined by normalization of the resulting expression levels to that of *thioredoxin* (*ZmTXN*, gene ID GRMZM2G066612) as described<sup>56</sup>. The primer sequences used in qRT-PCR analysis are in Table S3. All primer pairs were selected using the Primer3 program and were tested in 3% agarose gel to ensure the specificity and efficiency of each primer set.

**In Situ hybridization.** The template for preparing an *in situ* hybridization probe to localize lnc2653 was amplified from a maize genomic DNA sample with a primer pair: 5'-GAGCAAGTCCAGCTGTTACATATATCTTTATAGAATATATAATTATTGCTACTCTTACAT-3' and 5'-ACTATTCGGTGACCAGCTTGAGCACGGTTGTTGTTGTATATATTAATTGGATCTGACGCATCCC-3'. The DNA sample was cloned into the pGEM-T easy PCR cloning vector (Promega). Two types of RNA probes were prepared with DIG RNA Labeling Kit (Sigma-Aldrich) using the T7 and SP6 primers. These two RNA probes have sequences complementary to each other. Caryopses of maize 12 DAP inbred W22 were fixed, dehydrated, and embedded in Paraplast Plus paraffin (Sherwood Medical Co.) Sections were prepared from the embedded caryopses with a Leica RM2235 rotary microtome (Leica-Microsystems). Hybridization, washing, blocking, antibody labelling, and colour reaction were carried out as previously described<sup>12,57</sup>.

## References

- Lopes, M. A. & Larkins, B. A. Endosperm origin, development, and function. *Plant Cell* **5**, 1383–1399, doi:10.1105/tpc.5.10.1383 (1993).
- Olsen, O. A. ENDOSPERM DEVELOPMENT: Cellularization and Cell Fate Specification. *Annual review of plant physiology and plant molecular biology* **52**, 233–267, doi:10.1146/annurev-arplant.52.1.233 (2001).
- Olsen, O. A. Nuclear endosperm development in cereals and Arabidopsis thaliana. *Plant Cell* **16**(Suppl), S214–227, doi:10.1105/tpc.017111 (2004).
- Sabelli, P. A. & Larkins, B. A. The development of endosperm in grasses. *Plant Physiol* **149**, 14–26, doi:10.1104/pp.108.129437 (2009).
- Lafon-Placette, C. & Kohler, C. Embryo and endosperm, partners in seed development. *Curr Opin Plant Biol* **17**, 64–69, doi:10.1016/j.pbi.2013.11.008 (2014).
- Berger, F., Grini, P. E. & Schnittger, A. Endosperm: an integrator of seed growth and development. *Curr Opin Plant Biol* **9**, 664–670, doi:10.1016/j.pbi.2006.09.015 (2006).
- Becraft, P. W. & Yi, G. Regulation of aleurone development in cereal grains. *J Exp Bot* **62**, 1669–1675, doi:10.1093/jxb/erq372 (2011).
- Becraft, P. W. Endosperm imprinting: a child custody battle? *Curr Biol* **22**, R93–95, doi:10.1016/j.cub.2011.12.043 (2012).
- Sreenivasulu, N. & Wobus, U. Seed-development programs: a systems biology-based comparison between dicots and monocots. *Annu Rev Plant Biol* **64**, 189–217, doi:10.1146/annurev-arplant-050312-120215 (2013).
- Zhang, Z., Yang, J. & Wu, Y. Transcriptional Regulation of Zein Gene Expression in Maize through the Additive and Synergistic Action of opaque2, Prolamine-Box Binding Factor, and O2 Heterodimerizing Proteins. *Plant Cell* **27**, 1162–1172, doi:10.1105/tpc.15.00035 (2015).
- Gutierrez-Marcos, J. F. *et al.* maternally expressed gene1 Is a novel maize endosperm transfer cell-specific gene with a maternal parent-of-origin pattern of expression. *Plant Cell* **16**, 1288–1301, doi:10.1105/tpc.019778 (2004).
- Xiong, Y. *et al.* Adaptive expansion of the maize maternally expressed gene (Meg) family involves changes in expression patterns and protein secondary structures of its members. *BMC Plant Biol* **14**, 204, doi:10.1186/s12870-014-0204-8 (2014).
- Schnable, P. S. *et al.* The B73 maize genome: complexity, diversity, and dynamics. *Science* **326**, 1112–1115, doi:10.1126/science.1178534 (2009).
- Fatica, A. & Bozzoni, I. Long non-coding RNAs: new players in cell differentiation and development. *Nat Rev Genet* **15**, 7–21, doi:10.1038/nrg3606 (2014).
- Derrien, T. *et al.* The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression. *Genome Res* **22**, 1775–1789, doi:10.1101/gr.132159.111 (2012).
- Lu, X. *et al.* The differential transcription network between embryo and endosperm in the early developing maize seed. *Plant Physiol* **162**, 440–455, doi:10.1104/pp.113.214874 (2013).
- Li, L. *et al.* Genome-wide discovery and characterization of maize long non-coding RNAs. *Genome Biol* **15**, R40, doi:10.1186/gb-2014-15-2-r40 (2014).
- Wang, P. *et al.* Genome-wide high-resolution mapping of DNA methylation identifies epigenetic variation across embryo and endosperm in Maize (*Zea mays*). *BMC genomics* **16**, 21, doi:10.1186/s12864-014-1204-7 (2015).
- Zhang, W. *et al.* Genome-wide identification and characterization of maize expansin genes expressed in endosperm. *Mol Genet Genomics* **289**, 1061–1074, doi:10.1007/s00438-014-0867-8 (2014).
- Xin, M. *et al.* Dynamic parent-of-origin effects on small interfering RNA expression in the developing maize endosperm. *BMC Plant Biol* **14**, 192, doi:10.1186/s12870-014-0192-8 (2014).
- Chen, J. *et al.* Dynamic transcriptome landscape of maize embryo and endosperm development. *Plant Physiol* **166**, 252–264, doi:10.1104/pp.114.240689 (2014).
- Waters, A. J. *et al.* Comprehensive analysis of imprinted genes in maize reveals allelic variation for imprinting and limited conservation with other species. *Proc Natl Acad Sci USA* **110**, 19639–19644, doi:10.1073/pnas.1309182110 (2013).
- Zhang, M. *et al.* Genome-wide high resolution parental-specific DNA and histone methylation maps uncover patterns of imprinting regulation in maize. *Genome Res* **24**, 167–176, doi:10.1101/gr.155879.113 (2014).
- Xin, M. *et al.* Dynamic expression of imprinted genes associates with maternally controlled nutrient allocation during maize endosperm development. *Plant Cell* **25**, 3212–3227, doi:10.1105/tpc.113.115592 (2013).
- Makarevitch, I. *et al.* Genomic distribution of maize facultative heterochromatin marked by trimethylation of H3K27. *Plant Cell* **25**, 780–793, doi:10.1105/tpc.112.106427 (2013).
- Waters, A. J. *et al.* Parent-of-origin effects on gene expression and DNA methylation in the maize endosperm. *Plant Cell* **23**, 4221–4233, doi:10.1105/tpc.111.092668 (2011).

27. Trapnell, C., Pachter, L. & Salzberg, S. L. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* **25**, 1105–1111, doi:10.1093/bioinformatics/btp120 (2009).
28. Trapnell, C. *et al.* Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nature biotechnology* **28**, 511–515, doi:10.1038/nbt.1621 (2010).
29. Hueros, G., Varotto, S., Salamini, F. & Thompson, R. D. Molecular characterization of BET1, a gene expressed in the endosperm transfer cells of maize. *Plant Cell* **7**, 747–757, doi:10.1105/tpc.7.6.747 (1995).
30. Hueros, G., Royo, J., Maitz, M., Salamini, F. & Thompson, R. D. Evidence for factors regulating transfer cell-specific expression in maize endosperm. *Plant Mol Biol* **41**, 403–414 (1999).
31. Hueros, G. *et al.* Identification of a promoter sequence from the BETL1 gene cluster able to confer transfer-cell-specific expression in transgenic maize. *Plant Physiol* **121**, 1143–1152 (1999).
32. Haas, B. J. *et al.* De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat Protoc* **8**, 1494–1512, doi:10.1038/nprot.2013.084 (2013).
33. Storey, J. D. A direct approach to false discovery rates. *J Roy Stat Soc B* **64**, 479–498, doi:10.1111/1467-9868.00346 (2002).
34. Serna, A. *et al.* Maize endosperm secretes a novel antifungal protein into adjacent maternal tissue. *Plant J* **25**, 687–698 (2001).
35. Gruis, D. F., Guo, H., Selinger, D., Tian, Q. & Olsen, O. A. Surface position, not signaling from surrounding maternal tissues, specifies aleurone epidermal cell fate in maize. *Plant Physiol* **141**, 898–909, doi:10.1104/pp.106.080945 (2006).
36. Kohler, C. & Villar, C. B. Programming of gene expression by Polycomb group proteins. *Trends Cell Biol* **18**, 236–243, doi:10.1016/j.tcb.2008.02.005 (2008).
37. Lafos, M. *et al.* Dynamic regulation of H3K27 trimethylation during Arabidopsis differentiation. *PLoS Genet* **7**, e1002040, doi:10.1371/journal.pgen.1002040 (2011).
38. Quinodoz, S. & Guttman, M. Long noncoding RNAs: an emerging link between gene regulation and nuclear organization. *Trends Cell Biol* **24**, 651–663, doi:10.1016/j.tcb.2014.08.009 (2014).
39. Heo, J. B. & Sung, S. Vernalization-mediated epigenetic silencing by a long intronic noncoding RNA. *Science* **331**, 76–79, doi:10.1126/science.1197349 (2011).
40. Swiezewski, S., Liu, F., Magusin, A. & Dean, C. Cold-induced silencing by long antisense transcripts of an Arabidopsis Polycomb target. *Nature* **462**, 799–802, doi:10.1038/nature08618 (2009).
41. Gutschner, T. *et al.* The noncoding RNA MALAT1 is a critical regulator of the metastasis phenotype of lung cancer cells. *Cancer research* **73**, 1180–1189, doi:10.1158/0008-5472.CAN-12-2850 (2013).
42. Bailey, T. L. & Elkan, C. Fitting a mixture model by expectation maximization to discover motifs in biopolymers. *Proceedings/... International Conference on Intelligent Systems for Molecular Biology; ISMB. International Conference on Intelligent Systems for Molecular Biology* **2**, 28–36 (1994).
43. Gupta, M. Generalized hierarchical markov models for the discovery of length-constrained sequence features from genome tiling arrays. *Biometrics* **63**, 797–805, doi:10.1111/j.1541-0420.2007.00760.x (2007).
44. Zhan, J. *et al.* RNA sequencing of laser-capture microdissected compartments of the maize kernel identifies regulatory modules associated with endosperm cell differentiation. *Plant Cell* **27**, 513–531, doi:10.1105/tpc.114.135657 (2015).
45. Steffen, P. A. & Ringrose, L. What are memories made of? How Polycomb and Trithorax proteins mediate epigenetic memory. *Nat Rev Mol Cell Biol* **15**, 340–356, doi:10.1038/nrm3789 (2014).
46. Kang, B. H., Xiong, Y., Williams, D. S., Pozueta-Romero, D. & Chourey, P. S. Miniature1-encoded cell wall invertase is essential for assembly and function of wall-in-growth in the maize endosperm transfer cell. *Plant Physiol* **151**, 1366–1376, doi:10.1104/pp.109.142331 (2009).
47. Gehring, M. Genomic imprinting: insights from plants. *Annu Rev Genet* **47**, 187–208, doi:10.1146/annurev-genet-110711-155527 (2013).
48. Patten, M. M. *et al.* The evolution of genomic imprinting: theories, predictions and empirical tests. *Heredity* **113**, 119–128, doi:10.1038/hdy.2014.29 (2014).
49. Costa, L. M. *et al.* Maternal control of nutrient allocation in plant seeds by genomic imprinting. *Curr Biol* **22**, 160–165, doi:10.1016/j.cub.2011.11.059 (2012).
50. Ariel, F. *et al.* Noncoding transcription by alternative RNA polymerases dynamically regulates an auxin-driven chromatin loop. *Mol Cell* **55**, 383–396, doi:10.1016/j.molcel.2014.06.011 (2014).
51. Wang, X. *et al.* Genome-wide and organ-specific landscapes of epigenetic modifications and their relationships to mRNA and small RNA transcriptomes in maize. *Plant Cell* **21**, 1053–1069, doi:10.1105/tpc.109.065714 (2009).
52. Sekhon, R. S. *et al.* Genome-wide atlas of transcription during maize development. *Plant J* **66**, 553–563, doi:10.1111/j.1365-313X.2011.04527.x (2011).
53. Frith, M. C. *et al.* Discrimination of non-protein-coding transcripts from protein-coding mRNA. *RNA biology* **3**, 40–48 (2006).
54. Boerner, S. & McGinnis, K. M. Computational identification and functional predictions of long noncoding RNA in *Zea mays*. *PLoS One* **7**, e43047, doi:10.1371/journal.pone.0043047 (2012).
55. Kim, D. H., Zografos, B. R. & Sung, S. Mechanisms underlying vernalization-mediated VIN3 induction in Arabidopsis. *Plant Signal Behav* **5**, 13465 [pii], doi:10.1104/pp.110.161083 (2010).
56. Li, G. *et al.* Temporal patterns of gene expression in developing maize endosperm identified through transcriptome sequencing. *Proc Natl Acad Sci USA* **111**, 7582–7587, doi:10.1073/pnas.1406383111 (2014).
57. Wang, G. *et al.* An expression analysis of 57 transcription factors derived from ESTs of developing seeds in Maize (*Zea mays*). *Plant Cell Rep* **29**, 545–559, doi:10.1007/s00299-010-0843-7 (2010).

## Acknowledgements

This work was carried out with the support from the USDA NIFA Award AFRI grant (2011-67013-30119). This work was also supported by the grants from the Research Grants Council (RGC) of Hong Kong (GRF14126116 and AoE/M-05/12), Cooperative Research Program for Agriculture Science and Technology Development (Project No. 10953092015) Rural Development Administration, Republic of Korea, and by the NIH Grant (GM100108).

## Author Contributions

E.K., B.K., and S.S. designed experiments. Y.X. prepared RNA from cryo-dissected samples. E.K. analysed sequencing data. Y.P. generated qRT-PCR data. D.K. generated ChIP-qPCR data. E.K., B.K., and S.S. wrote the manuscript.

## Additional Information

**Supplementary information** accompanies this paper at doi:10.1038/s41598-017-03878-4

**Competing Interests:** The authors declare that they have no competing interests.

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2017