Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Spatial and single-nucleus transcriptomics reveal the complexity of genomic imprinting in maize

Tao Li
The Shenzhen Campus of Sun Yat-sen University, Sun Yat-sen University
Yi Jiang
The Shenzhen Campus of Sun Yat-sen University, Sun Yat-sen University
Mingyue Zhang
Shandong Agricultural University
Kesen Zhu
The Shenzhen Campus of Sun Yat-sen University, Sun Yat-sen University
Siqi Jiang
The Shenzhen Campus of Sun Yat-sen University, Sun Yat-sen University
Xuerong Yang
Shandong Agricultural University
Hongjun Liu
Shandong Agricultural University
Jiechen Wang
Center for Excellence in Molecular Plant Sciences
Xiaomei Dong
Shenyang Agricultural University
Junpeng Shi
shijp6@mail.sysu.edu.cn

The Shenzhen Campus of Sun Yat-sen University, Sun Yat-sen University

Research Article

Keywords: Maize, spatial transcriptomics, genomic imprinting, allele-specific expression, lncRNAs

Posted Date: June 26th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10052204/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Version of Record: A version of this preprint was published at Genome Biology on September 28th, 2026. See the published version at https://doi.org/10.1186/s13059-026-04291-9.

Spatial and single-nucleus transcriptomics reveal the complexity of genomic imprinting in maize

Tao Li[1], Yi Jiang[1], Mingyue Zhang[2], Kesen Zhu[1], Siqi Jiang[1], Xuerong Yang[2], Hongjun Liu[2,3], Jiechen Wang[4], Xiaomei Dong[5] & Junpeng Shi[1,#]

[1]Guangdong Provincial Key Laboratory of Plant Stress Biology, School of Agriculture and Biotechnology, The Shenzhen Campus of Sun Yat-sen University, Sun Yat-sen University, Shenzhen, 518107, China

[2]State Key Laboratory of Crop Biology, College of Life Sciences, Shandong Agricultural University, Tai'an, Shandong, 271018, China

[3]Yazhouwan National Laboratory, Sanya, Hainan, 572024, China

[4]National Key Laboratory of Plant Molecular Genetics, CAS Center for Excellence in Molecular Plant Sciences, Shanghai Institute of Plant Physiology & Ecology, Shanghai, 200032, China

[5]College of Bioscience and Biotechnology, Shenyang Agricultural University, Shenyang, Liaoning, 110866, China

[#]Correspondence: Junpeng Shi (shijp6@mail.sysu.edu.cn)

Keywords: Maize, spatial transcriptomics, genomic imprinting, allele-specific expression, lncRNAs

Abstract

Genomic imprinting is a classical epigenetic phenomenon first discovered in maize, which occurs almost exclusively in the triploid endosperm and is thought to mediate parental conflicts over offspring survival. Although previous studies using manually dissected maize endosperm from reciprocal hybrids coupled with RNA sequencing have uncovered hundreds of putative imprinted genes, the intensive cellular heterogeneity within endosperm has been overlooked, leaving it unclear whether imprinting varies across different cell types. Here, we applied 10x spatial transcriptomics and single-nucleus RNA sequencing (snRNA-seq) to profile reciprocal maize kernels, resolving six major endosperm cell types. We reveal a high degree of cell-type specificity of allele-specific expression (ASE) genes that approximately two-thirds of them occurring in a single endosperm cell type. Additionally, over 70% of imprinted genes identified by spatial transcriptomics can be validated by snRNA-seq at the cell-type level. Fewer than half of them are found in multiple endosperm cell types, and paternally expressed genes (PEGs) outnumber maternally expressed genes (MEGs) by approximately two-fold. We disclosed an imprinted long non-coding RNA (lncRNA) that clustered with the typical MEG *Mez1* in a maternally hypomethylated region. Our findings provide fresh evidence that post-fertilization endosperm differentiation may shape the spatial and cellular complexity of genomic imprinting in maize.

Introduction

The maize (*Zea mays ssp. mays*) kernel stands as a preeminent model system in plant genetic studies and a critical organ of global agricultural significance[1-3]. Comprising a diploid embryo, a triploid endosperm, and the surrounding maternal seed coat, it has been the foundation for seminal genetic discoveries. Notably, Barbara McClintock's identification of transposable elements stemmed from observations of mosaic pigmentation on maize kernels[4]. Intriguingly, it was also within the context of non-Mendel inheritance of endosperm color, governed by the *R* gene, that genomic imprinting (an epigenetic phenomenon dictating parent-of-origin-specific gene expression) was first described[5]. Imprinted genes are classified as PEGs and MEGs[6]. Classical forward and reverse genetic approaches in maize and other plants have identified some well-known imprinted genes, such as *Meg1*, which is specifically expressed in the basal endosperm transfer layer (BETL) and facilitates nutrient transport from maternal tissue to the developing kernel[7,8]. Furthermore, the *Fie* (fertilization independent endosperm), a component of polycomb repressive complex 2 that initially characterized as MEG in *Arabidopsis thaliana*, exhibits canonical imprinting in both rice (*OsFIE1*[9-11]) and maize (*Fie1*[12-14]), underscoring the deep evolutionary conservation of this epigenetic mechanism. Transcriptome-wide identification of gene imprinting has been revolutionized by RNA sequencing of endosperm from reciprocal hybrids, leveraging parental DNA polymorphisms to ascertain ASE. Such studies across diverse maize genotypes and developmental stages

typically identify 100-200 imprinted genes, revealing context-dependent variation[15-19]. Critically, these analyses have universally treated the endosperm as a homogeneous tissue, neglecting its profound spatial organization into distinct functional domains[1,20,21], including the aleurone layer (AL), BETL, endosperm adjacent to scutellum (EAS), conducting zone (CZ), starchy endosperm (SE), vitreous endosperm (VE), as well as inner cellular heterogeneity within each compartment. Since genomic imprinting can be influenced by both primary imprints established prior to fertilization and secondary imprints established during cellular differentiation post-fertilization[22], it should have some degree of cellular heterogeneity, whereas the spatial and cell-type specific patterning of imprinted genes within plant endosperm remain largely unexplored territory.

Single-cell sequencing (scRNA-seq) or snRNA-seq and spatial transcriptomics have emerged as transformative tools for deconvoluting cellular complexity within tissues. While pioneering studies have applied scRNA-seq to investigate maize kernel and endosperm, their use of genetically uniform inbred lines (e.g., B73, W22) inherently precludes the discrimination of parental alleles, a prerequisite for imprinting analysis[21,23]. By comparison, mammalian systems benefit from natural genetic heterozygosity, enabling single-cell resolution studies of ASE or imprinting. These studies have revealed striking cellular specificity within human tissues[24-27], mouse brain[28], and notably the porcine placenta[29], where ~70% of imprinted genes were restricted to a single cell type. A recent study in *Arabidopsis* employed low-input Smart-

Seq2 on ~800 nuclei isolated from endosperm of reciprocal Col × Cvi crosses, successfully demonstrating cell-type specific imprinting, with the CZ exhibiting a particularly pronounced effect[30]. Applying these cutting-edge snRNA-seq and spatial transcriptomics technologies to developing kernels from genetically defined reciprocal crosses presents an unprecedented opportunity to dissect ASE, thereby enabling the interrogation of genomic imprinting across various endosperm functional domains and cell types. Deciphering this spatiotemporal regulatory landscape in maize—the very organism where imprinting was first discovered—promises not only to illuminate fundamental mechanisms of epigenetic regulation in plant development but also to reveal novel dimensions of a phenomenon central to seed viability and yield.

Here, we generate a spatial transcriptomic profile of the longitudinal section per reciprocal hybrid maize kernels along with matched snRNA-seq data to explore the cell-type specificity of imprinted genes. Among six major endosperm cell types, we find about two-thirds of ASE genes occurred in one cell type. Both spatial transcriptomics and snRNA-seq data identify approximately twice as many PEGs as MEGs, while fewer than half of the imprinted genes exhibited imprinting in multiple endosperm cell types. We unlocked a maternally expressed lncRNA that clusters with the canonical MEG *Mez1*, and demonstrate that parental transcriptional programs sequentially diverged along the developmental trajectory of maize endosperm. Our study provides proof-of-principle evidence that post-fertilization cellular differentiation shapes the spatial and cellular complexity of genomic imprinting in maize, and offers a

promising set of candidate genes for functional dissection of endosperm development.

Results

Spatial and single-nucleus transcriptomes of reciprocal maize kernels

To generate the transcriptome of hybrid maize kernels, we performed reciprocal crosses between two maize inbred lines B73[31] and Mo17[32], and collected their developing kernels at 14 days after pollination (DAP). At this developmental stage, the maize kernels have completed the differentiation of major cell types, endosperm cells have ceased proliferation, and storage metabolites commence synthesis and accumulation[21]. We performed spatial transcriptome sequencing on the maximal longitudinal plane of reciprocal kernels (BM, B73 × Mo17; MB, Mo17 × B73) using 10x Genomics Visium technology. Meanwhile, the hand-dissected endosperm of BM and MB at same developmental stage were sequenced by DNBelab C4 snRNA-seq technology with two replicates per sample (Fig. 1a). The Visium sections of BM and MB exhibited highly similar anatomical structure under hematoxylin and eosin staining (Fig. 1b, c), with the small-sized embryo (EM) region exhibited distinctly higher tissue density and RNA abundance than the endosperm region (EN, Fig. S1). We used bbknn[33] to remove batch effects, then merged all data spots (3,216 from BM and 3,618 from MB) for downstream analysis after quality control. The mean and median gene counts per spot were 5,192 and 4,965, respectively, and the number of analyzable genes in both BM and MB approached saturation as the number of spots increased (Fig. S2). For snRNA-seq, a

total of 38,066 nuclei (BM, n = 18,554; MB, n = 19,512) passed quality control, and nuclei from reciprocal crosses and replicates exhibited similar space distributions when integrated in a single UMAP (Fig. S3), indicating the high quality of our data (Supplementary Table 1).

We then performed dimensionality reduction and unsupervised clustering on spatial RNA-seq data, identifying 21 distinct clusters each had 142 to 568 spots (Fig. 1d). UMAP projection of all spots revealed that while reciprocal crosses introduced a degree of variation in expression profiles, the substantial co-localization of BM and MB spots within the same clusters indicated that the expression profiles defining distinct cell clusters were largely conserved between reciprocal crosses (Fig. S4, S5). Using a set of previously reported marker genes[21], we further annotated our data into nine functional compartments (Fig. 1e, f). These compartments, ranked by spot numbers, were VE, pericarp (PE), CZ, EM, placento-chalazal (PC), EAS, BETL, SE and AL (Fig. S6). All compartments were identified in both BM and MB, whereas MB exhibited minimal number of spots annotated as SE. This difference may be attributed to developmental or anatomical variations between reciprocal crosses (Fig. S7) and/or unavoidable spatial discrepancies during tissue sectioning. We further identified 2,564 differentially expressed genes (DEGs) among nine compartments, among which approximately 72% came from EM, highlighting the pronounced divergence of EM relative to other endosperm compartments (Fig. S8 and Supplementary Table 2). Among these DEGs, 98 encoded transcription factors (TFs). This TF group included

144 well-characterized regulators of endosperm development, such as *O2*, *Pbf1*, and *Nkd1*,
145 which exhibited the expected significantly higher expression in VE. *O11* encoded a
146 bHLH family TF[34]; however, unlike *O2* which was primarily expressed in the upper
147 endosperm regions (VE and SE), *O11* expression was more spatially restricted to the
148 basal endosperm domains, including EAS, BETL, and CZ. We also identified 12 TFs
149 (including *Sbp17*, *Hb118*, and *Mybr29*) highly expressed in BETL, a highly specialized
150 cell type that functions in nutrient transport to nourish the developing kernel.
151 Intriguingly, we found several MADS-box TFs associated with floral organ
152 development (*Zag2*, *Bde1* and *Zmm1*) that highly expressed in the PC, as well as *Zmm6*,
153 *Zmm7* and *Zmm27* that expressed highly in both PC and PE (Fig. 1g). Their active
154 expression in maternal seed coat tissues might be inherited from their prior transcription
155 in the inflorescence meristem[35]. Patterns of spatial expression of above-mentioned TFs
156 and well-established marker genes could be visualized and validated by the electronic
157 RNA *in situ* hybridization in both BM and MB spatial transcriptome maps (Fig. S9,
158 S10). Finally, we processed the snRNA-seq data and identified 20 clusters via
159 unsupervised clustering, which were also annotated into six distinct endosperm cell
160 types (Fig. 1h) that could enable downstream comparison with the cell types defined by
161 spatial RNA-seq data.
162
163 **ASE genes exhibit high degree of cellular specificity in maize endosperm**
164 To further dissect the contribution of parental alleles to gene expression in spatial

transcriptome, we developed a pipeline called spaceASE (Fig. S11) which calculated the proportion of maternal reads (PMR, representing the B73 allele in BM and Mo17 allele in MB) in each spot (Fig. 2a and Fig. S12). PMR were consistently near 1.0 in maternal tissues PC and PE, ~0.5 in EM, whereas most EN spots showed PMR between 0.6 and 0.7, confirming the high genotyping accuracy of spaceASE. We then performed ASE identification across seven offspring cell types (one in EM and six in EN) in both BM and MB. Except for the SE in MB due to its low spot count for reliable comparison, we found between 106 (AL in MB) and 590 (VE in MB) ASE genes in remaining 13 cell types (Fig. 2b, c and Supplementary Table 3). Notably, a consistently higher number of ASE genes favoring B73 expression was found for all cell types, with only one exception of AL in MB (Fig. 2b, c). Integrating ASE data from BM and MB allowed us to distinguish between sequence-dependent ASEs (sdASEs, which refer to genes that consistently favor the same allele in both reciprocal crosses) and imprinted genes (genes show a parent-of-origin effect on gene expression). A total of 229 B73-sdASEs and 133 Mo17-sdASEs have been identified, with the maximal and minimal numbers of sdASEs occurred in VE and AL (Fig. 2d, e and Fig. S13), respectively. It was worth noting that the majority of ASEs could not be classified as either sdASE or imprinted genes (Fig. 2b, c), implicating the stochastic effects as a predominant driver for most ASE genes.

To explore the cellular specificity of ASEs, we then focused our analysis on endosperm cell types (VE, BETL, AL, CZ, EAS and SE), which were all derived from a single central cell fertilized by the sperm. This analysis revealed that ASE genes

exhibit a high degree of cell-type specificity, with approximately 59.4%–68.1% of genes detected in only a single cell type (Fig. 2f and Fig. S14). We also identified ASE genes in six endosperm cell-types in snRNA-seq data, which displayed highly similar results that 59.9%–70.1% of ASE genes were identified in a single cell type (Fig. S15). In both BM and MB, VE exhibited highest number of cell type specific ASE genes, which might be associated with its largest number of spots and genes that could be profiled (Fig. 2f and Fig. S14). Notably, a greater proportion (over 70%) of sdASE genes were found in only one cell type, further supporting the strong cellular specificity of ASE genes. Intriguingly, we identified three genes exhibiting antagonistic ASE states across distinct endosperm cell types. As an example in BM, *Zm00001eb026680*, which encoded a heat shock protein, displayed B73-ASE in BETL and CZ while Mo17-ASE in SE (Fig. S16). Even within a single cell type, pronounced heterogeneity in ASE was evident across individual spots. For instance, among 477 analyzable spots in VE for *Plt43* (a gene involved in phospholipid transfer), 19.07% displayed B73-ASE, 6.30% exhibited a biallelic expression pattern, and the remaining 74.63% showed Mo17-ASE (Fig. 2g, h). A similar mosaic-like ASE was also observed for *Umc161a* (Fig. 2i), a gene encoding Gly-Tyr-Phe domain and actively expressed in developing seeds. These results indicate that ASE genes exhibit a high degree of cellular specificity across different compartments of maize endosperm.

Cell-type resolved imprinted genes in maize endosperm

Our spatially resolved transcriptome data from reciprocal crosses enabled the analysis of imprinted genes across distinct endosperm cell types. We defined MEGs as genes showing B73-ASE in BM and Mo17-ASE in MB simultaneously; conversely, PEGs were those that showing Mo17-ASE in BM and B73-ASE in MB (Fig. 3a). We initially identified MEGs and PEGs in each of the six endosperm cell types, yielding a total of 49 MEGs and 56 PEGs (Fig. S17). Given our observation of maternal RNA influence on the AL and BETL, we further filtered MEGs in these cell types by excluding those highly expressed in adjacent maternal cells. This stringent filtering resulted in a final set of 29 MEGs and 56 PEGs (Fig. 3a and Supplementary Table 4). The overall finding that PEGs outnumbered MEGs is consistent with previous reports using manually dissected endosperm coupled with short-read RNA sequencing[16]. Functional enrichment analysis revealed distinct roles for MEGs and PEGs: MEGs were significantly enriched for processes related to epigenetic regulation and embryonic development, while PEGs were mainly involved in signaling pathways, metabolic regulation, and mRNA processing (Fig. S18). We also generated the whole-genome bisulfite sequencing data of BM and MB endosperm, and observed higher CG and CHG methylation level on the paternal allele over the maternal allele near the transcription start site of MEGs (Fig. S19).

Using the same pipeline and criteria described above, we identified a similar number of imprinted genes (28 MEGs and 47 PEGs) in the snRNA-seq data, and further used these genes to validate the imprinted genes from spatial RNA-seq data (Fig. 3b, c

and Supplementary Table 5). For MEGs, except for the SE cell type in which only one MEG was identified, 67% to 80% of analyzable MEGs in other five cell types could be validated by the snRNA-seq data. PEGs showed higher validation ratio over MEGs, exceeding 80% in AL, BETL, SE, and CZ (Fig. 3c). Notably, spatial imprinted genes that could not be validate by snRNA-seq data still often exhibited pronounced parental bias in the snRNA-seq data. When we weakened the criteria of parental bias for calling imprinted genes by 10% (from 90% to 80%), the overall recall ratios for MEGs and PEGs increased to 88% and 92%, respectively (Fig. 3b). These results indicate the overall high-confidence of cell-type resolved imprinted genes identified by spatial transcriptomic data.

We next investigated the cell-type specificity of previously reported canonical imprinted genes in spatial transcriptomic data, including *Fie1*, *Fie2*, *Nrp1*, *Mez1*, *Meg1*, *Mee1*, *Peg1*, and *Ded1*[15,36,37]. Among these, no homologous sequence was identified for *Peg1* in B73_V5 reference genome, the locus homologous to *Mee1* in B73 was not annotated as a gene, and *Meg1* showed identical DNA sequence between B73 and Mo17, precluding their imprinting analysis (Fig. S20). For other reported MEGs, we found that *Fie1*, *Nrp1* and *Mez1* consistently exhibited maternal expression and were conserved across most endosperm cell types (Fig. 3d, f and Fig. S21a-d). In contrast, *Fie2* showed lower expression level than *Fie1*, and displayed a mosaic pattern of endosperm spots favoring maternal and paternal expression (Fig. S21e, f), suggesting divergent imprinting regulation of these two paralogs after maize tetraploidization. Although our

data showed a consistent paternal expression bias for the reported PEG *Ded1* (Fig. S22), it did not meet our stringent criteria requiring >= 90% paternal reads in any one of the six endosperm cell types, thus not classified as a PEG.

To further quantify the cell-type specificity of imprinted genes, we first focused on the more accurately identified PEGs. Among all 56 PEGs, 34 were detected in only one cell type (27 in VE, 4 in BETL, and 3 in EAS). This proportion (61%) is similar to the aforementioned ratio of cell-type specific ASE genes. We further examined the imprinting status of these PEGs in cell types other than the one that imprinting happened. There were two scenarios that distinguished these cell-type specific PEGs: First, a gene may be expressed in multiple cell types but imprinted in only one of them (i.e., its imprinting status can be assessed across several cell types, yet imprinting is observed in only one). Second, a gene may be expressed exclusively in one cell type and imprinted specifically in that same expressing cell type. We found 24 PEGs fall into the first category, suggesting genuine complexity of imprinting, and 10 PEGs belong to the second (Fig. S23). Most of cell-type specific PEGs have not been previously reported as imprinted genes, including *Trps8* (*trehalose-6-phosphate synthase 8*) in BETL, *Nrg1* (*nitrate regulatory gene 1*) in EAS, as well as *Thx41* (*trihelix transcription factor 41*) and *Sbp3* in VE. Notably, the *Ippi1* (*isopentenyl pyrophosphate isomerase 1*) gene exhibited PEG status across all six endosperm cell types (Fig. 3e, g), highlighting its potential importance in endosperm development that deserved subsequent investigation. We also analyzed the cell-type specificity of 29 MEGs, among which 12 (41.38%) were

identified as cell-type specific MEGs in AL or VE. Notable examples of cell-type specific MEGs included *Zmm7* in VE, whereas well-known MEGs such as *Fie1*, *Nrp1*, and *Mez1*, as well as curated genes like *Pyl2* and *Ubi3*, displayed maternal expression status in multiple cell types (Fig. S23).

Imprinted genes are rare in the maize embryo

Except the well-documented imprinted genes in maize endosperm, its existence in maize embryo has long been a subject of debate[22,38-40]. Our data enabled the analysis of genomic imprinting in endosperm and embryo simultaneously, which resulted in six potential MEGs, but no PEGs, in the embryo (Fig. S24). In contrast with the canonical phenomenon of higher PEG number in endosperm, we suspected that MEGs in embryo might be influenced by maternal tissues or adjacent endosperm cells, thus we analyzed their expression and imprinting status in embryo-adjacent endosperm cell types: AL, BETL and SE, clarifying that these genes were also MEGs in at least one of these adjacent cell types (Fig. S24), suggesting that the embryo lacks specific imprinted genes distinct from those existed in the endosperm. To offer more clue into these embryo MEGs, we analyzed their PMR across individual spots. Spots supporting MEG status (exhibiting high PMR) in embryo frequently displayed a mosaic-like distribution (Fig. S24), indicating these genes were not consistently imprinted across cells in the embryo. Collectively, our data suggest that genomic imprinting rarely happened in maize embryo.

Spatial profiling of imprinted lncRNAs

While the existence of imprinted lncRNAs had been reported in both maize and mammals[15,41], their spatial distribution and cell heterogeneity within maize kernels remain largely unexplored. To build a comprehensive lncRNA dataset in maize kernels, we annotated 7,831 intergenic lncRNAs, comprising 13,084 isoforms (Fig. 4a and Fig. S25a), using long-read transcriptome data for the embryo and endosperm of BM and MB. These lncRNAs exhibited comparable transcript length to protein-coding genes (Fig. S25b), while they possessed fewer exons, with individual exons being significantly longer than those of protein-coding genes (Fig. S25c). We further quantified the expression of these lncRNAs across nine cell types in our spatial transcriptome data. Globally, lncRNAs displayed substantially higher cell-type specificity than protein-coding genes: 76% of lncRNAs were detected in only one cell type, markedly exceeding the 50% observed for protein-coding genes (Fig. 4b and Fig. S25d). Categorizing lncRNAs as cell-type specific, intermediate or ubiquitous based on their expression patterns revealed a negative correlation between expression level and cell-type specificity. Protein-coding genes exhibited a similar feature, yet their differences between categories were much smaller (Fig. 4c).

We further identified 169 lncRNAs exhibiting ASE across seven embryo and endosperm cell types. In contrast to the pronounced B73-ASE observed for protein-coding genes, the ASE of lncRNAs showed no significant bias towards B73, especially

312 in MB that the ASE of lncRNAs was more frequently found for Mo17 allele (Fig. 4d,
313 e). Over 80% of lncRNAs exhibited ASE originating from VE, reflecting the
314 requirement of sufficient read depth to statistically assess parental contributions and
315 call ASE lncRNAs, since the data depth of lncRNAs was much lower in cell types other
316 than VE (Fig. S26).

317 We next explored imprinted lncRNAs and identified three strictly imprinted
318 candidates (*STRG.17426*, *STRG.27146*, and *STRG.38589*) located on chromosomes 4,
319 6, and 9, respectively (Fig. S27). Among them, *STRG.27146* is a single exon lncRNA
320 approximately 4,000 bp, situated ~4,700 bp upstream of *Mez1*. To validate its transcript
321 structure, we visualized the read alignment in its associated genomic region using IGV.
322 Multiple full-length non-chimeric (FLNC) reads spanned this lncRNA, supporting the
323 authenticity of the assembled transcript (Fig. 4f). In addition, CPC2[42] prediction
324 indicated that this transcript possesses low coding potential. Together, these results
325 suggest that this transcript is a *bona fide* lncRNA. Spatial PMR analysis of *STRG.27146*
326 suggested a maternally expressed status that consistent with its associated MEG *Mez1*
327 (Fig. 4g). The maternal-biased expression of this lncRNA could be independently
328 validated in most single-nuclei of the snRNA-seq data (Fig. 4h). Using the whole-
329 genome bisulfite sequencing data, we could identify the hypomethylation (CG and
330 CHG) of maternal allele spanning this lncRNA-*Mez1* cluster compared with the
331 paternal allele in the endosperm of both BM and MB (Fig. 4i), whereas both the
332 maternal and paternal alleles maintained substantially high methylation level in embryo

(Fig. S28). The other two imprinted lncRNAs also displayed strong maternal expression patterns and could be validated by snRNA-seq data, with *STRG.38589* exhibiting relatively lower expression level (Fig. 4j and Fig. S29, S30). In summary, we characterized the strong cell-type heterogeneity of lncRNA expression and disclosed a lncRNA-gene imprinting cluster that showing the hypomethylation of maternal allele in maize endosperm.

Haplotype-resolved pseudotime analysis of maize endosperm by snRNA-seq

Our reciprocal-hybrid snRNA-seq data provided an opportunity to dissect parental contributions during maize endosperm development at single-nucleus resolution. We separately constructed the pseudotime trajectories of BM and MB nuclei using Monocle2[43], which yielded highly similar topologies between reciprocal crosses. Specifically, nuclei from AL positioned at one branch tip, while SE nuclei situated at the tips of other two major branches, with VE broadly distributed across the trajectory tree (Fig. 5a, b and Fig. S31a, b). According to the prior report that AL is less differentiated than SE[23], we could root the trajectory starting from AL to developmental progression. We then computed the maternal and paternal pseudotime independently for each nucleus and projecting the parental difference onto the trajectory (see Methods). Parental pseudotimes were largely concordant in the trajectory branch defining as AL and in a subset of CZ and BETL cells, but diverged strongly toward the SE and EAS branch tips (Fig. 5c and Fig. S31c). These findings suggest a parental asymmetry during

the endosperm differentiation of maize hybrids that the heterogeneity between parents progressively increasing as development proceeds.

Discussion

The endosperm, a highly complex organ in higher plants, is typically triploid and originates from the fertilization of the central cell by a sperm cell[2,20]. Post-fertilization endosperm differentiation is accompanied by highly dynamic gene expression variation, which occurs not only among cells but also between parental alleles, leading to the cellular complexity of ASE and genomic imprinting[44-47]. Our study leverages spatial transcriptomics together with matched single-nucleus transcriptomics in reciprocal maize hybrids kernel to investigate this complexity, providing proof-of-concept evidence for dissecting allelic complexity at spatial and single-nucleus resolution. We unveil a remarkable spatial specificity of ASE, with the majority of ASE genes being confined to a single endosperm cell type. This cellular specificity, coupled with the discovery of genes exhibiting antagonistic ASE in different cell types and mosaic expression within a single type, highlights a dynamic and context-dependent regulatory landscape. By applying stringent filtering, we refined the catalogue of imprinted genes and examined them across cell types, confirming a global predominance of PEG over MEG but also suggesting that PEGs exhibit greater cell-type specificity[16]. Crucially, our data suggests that genuine embryo-specific imprinting is rare in maize, as putative embryonic MEGs were either shared with adjacent endosperm tissues or displayed

inconsistent mosaic patterns. Furthermore, we extended this spatial analysis to lncRNAs, characterizing their intensive cell-type specificity and identifying a novel imprinted lncRNA-mRNA cluster coordinated by maternal allele hypomethylation.

It is important to note that our spatial transcriptomic data represent a single, thin section of the maize kernel and thus cannot fully capture the complete repertoire of imprinted genes and their heterogeneity within the intact 3D organization of maize seeds. A potential solution is to collect serial sections of the kernel to generate spatial transcriptome data from multiple layers, as had been done by many spatial transcriptome studies studying brain development in primates, then integrating them into a unified 3D model[48]. In addition, SNP-based approaches and the use of a single reference genome carry certain limitations, and this choice often involves an important trade-off in the course of a study. First, a prerequisite for a gene to be analyzable is that its 3′ end contains high-quality SNPs, which is directly influenced by the size of the SNP set. A larger SNP set tends to result in a higher false-positive rate, whereas a smaller SNP set reduces the number of analyzable genes. Therefore, the size of the SNP set included in the analysis is a critical decision when investigating ASE and imprinting, and should be determined according to the study objectives. Second, the choice between using a single reference genome and a merged parental genome also entails a trade-off. A single genome cannot simultaneously account for presence/absence variation, copy-number variation, and structural variation between the two parental haplotypes, meaning that certain loci cannot be reliably aligned and are thus excluded from the

analysis. The use of merged genomes, in turn, requires that high-quality reference assemblies already exist for both parents, and inconsistencies in assembly or annotation quality between the parental genomes can introduce bias. Overall, the research community has not yet reached a consensus on analytical methods for ASE and imprinting, and the relative applicability of different approaches remains to be systematically evaluated.

Our work opens a new avenue for future research and validates the utility of spatial and single-nucleus transcriptomics for the study of ASE and genomic imprinting. Looking forward, integrating spatial and single-nucleus transcriptomics with other complementary technologies will further advance our understanding of ASE and imprinting. For instance, spatially or single-cell resolved profiling of DNA methylation and chromatin accessibility will be essential to elucidate how chromatin states dictate parental expression preferences in a cell-type-specific manner[49-52]. Coupling long-read transcriptome sequencing with single-cell or spatial approaches would more effectively resolve expression differences between parental alleles and reveal distinct imprinted isoforms[53]. The functional significance of the newly identified cell-type-specific imprinted genes, such as *Trps8* in the BETL and *Thx41* in the VE, warrants in-depth investigation through genetic and biochemical approaches to define their roles in nutrient transport, metabolic regulation, and endosperm patterning.

In summary, our study advances the understanding of genomic imprinting at higher resolution, extending it from the bulk-tissue level to the cell-type level, and uncovers a

set of cell-type-defined gene resources. These findings provide a valuable resource for subsequent gene identification and offer new insights into kernel development.

Methods

Plant growth and sample collection

The maize inbred lines were planted in the experimental station of Shandong Agricultural University (Tai'an, Shandong Province) in 2024. At the appropriate pollination stage, reciprocal crosses between B73 and Mo17 were performed to obtain hybrid kernels. Well-developed kernels from the middle portion of the ear were collected at 14 DAP and used to construct 10x Genomics Visium spatial transcriptome libraries. Endosperm was manually dissected for snRNA-seq library construction.

10x Genomics Visium library preparation and sequencing

Kernels for spatial transcriptome sequencing were embedded in OCT compound and snap-frozen on dry ice. Spatial transcriptome libraries were prepared using the 10x Genomics Visium protocol. Embedded blocks were cryosectioned, and structurally intact frozen sections were mounted onto 10x Visium Spatial Gene Expression slides. Sections were then fixed with formaldehyde, stained with HE, and imaged using a slide scanner. Notably, the effective capture area of the 10x Genomics Visium slide is a 6.5 × 6.5 mm square containing 4,992 spots, each with a diameter of 55 μm, which is sufficient to cover and capture the spatial transcriptomic landscape of the largest transverse section of the maize kernel. At 14 DAP, maize endosperm cells have an average diameter of approximately 60 μm, whereas embryo cells range from 10 to 30 μm in diameter[21]. Consequently, a single spot may capture RNA molecules originating

from one or a few cells. This multi-cellular mixing effect is expected to be more pronounced in the embryo, where cells are more densely packed, resulting in expression profiles that reflect a composite transcriptional signature of multiple neighboring cells. In addition, cell mixing may also occur along the vertical axis, depending on the thickness of the tissue section and whether the section intersects multiple vertically stacked cells. In this study, tissue sections were prepared at a thickness of 10 μm, which is smaller than the diameter of most kernel cells. Therefore, vertical mixing effects are primarily attributable to cases in which a section simultaneously intersects two overlapping cells. Overall, we anticipate that each spot contains a certain degree of cellular mixing and represents the aggregated transcriptional profile of a local group or cluster of spatially adjacent cells with similar gene expression characteristics. Tissue permeabilization, reverse transcription, cDNA amplification, and library construction were performed following the manufacturer's protocol using the Visium Spatial Gene Expression Slide Kit (PN-1000185). Spatial transcriptome libraries were sequenced on Illumina platform by Shanghai OE Biotech.

snRNA-seq library preparation and sequencing

snRNA-seq was performed on the DNBelab C4 droplet microfluidics platform. Nuclei were first isolated from BM and MB endosperm and prepared as a single-nucleus suspension, which was subjected to quality control with the requirements that nuclei be intact and free of visible damage, and that the nuclear concentration and suspension

background meet the loading criteria, with no obvious debris or clumping. Qualified samples were loaded according to the target nucleus number. Water-in-oil droplets were used to co-encapsulate single nuclei with barcoded capture beads using a dual-bead capture system, in which Cell Beads captured the polyA structure of mRNA and Index Carrier beads identified and merged the Cell Beads based on the strong correlation among similar barcode primers within the same droplet. A reverse transcription reaction system was assembled within the droplets, after which the emulsion was broken to recover the products, followed by cDNA amplification, purification, and library quality control. Libraries that passed quality control were sequenced on the DNBSEQ-T7 platform.

Whole genome bisulfite sequencing (WGBS)

Genomic DNA was first spiked with unmethylated lambda DNA (serving as a negative control) and fragmented using a Bioruptor sonicator (Diagenode). The fragmented DNA was then subjected to end repair, ligation with methylated adapters, magnetic bead-based cleanup, and fragment size selection. The selected fragments were amplified by PCR and purified to generate the final bisulfite sequencing libraries. Library quality and concentration were assessed using Qubit 2.0 Fluorometer (Life Technologies, CA, USA) and Agilent Bioanalyzer 2100 system. Qualified libraries were sequenced on the Illumina NovaSeq platform using the paired-end 150 bp mode.

Processing and annotation of 10x Genomics Visium data

Raw Visium spatial transcriptome reads were aligned to the B73_RefGen_v5 reference genome using the Space Ranger (v2.1.1) to obtain filtered count matrices. Downstream analysis was performed in Scanpy (v1.11.0) [54]. Quality control was applied with the following criteria: (1) each spot had at least 500 genes and a total count of no less than 1,000 UMIs; (2) each gene had to be expressed in at least 3 spots. Filtered gene expression matrices were normalized and batch corrected. Gene counts for each spot were scaled to 1,000,000 and log-transformed using normalize_total and log1p functions in Scanpy. The top 2,000 highly variable genes were selected for principal component analysis, and batch effects were removed using the bbknn method[33]. Unsupervised clustering was performed using the Leiden algorithm (resolution = 1.2), and t-SNE and UMAP (min_dist = 0.4) were used for visualization and cluster exploration. A curated list of spatially resolved marker genes from Fu et al. [21] was converted to B73_RefGen_v5 coordinates and used to train a CellAssign annotation model. Training was performed on an NVIDIA A800 GPU with max_epochs = 300[55,56]. Preliminary spot annotations were assigned using the trained CellAssign model, and final tissue-type annotations were determined by integrating assignment probabilities, marker gene expression, and anatomical features.

Processing and annotation of snRNA-seq data

snRNA-seq raw reads were aligned to the B73_RefGen_v5 reference genome using

dnbc4tools (v3.0) to generate a filtered gene-expression count matrix. Downstream analysis was performed in Scanpy (v1.11.0). First, the number of nuclei to retain was determined by knee-point evaluation based on the target number of captured nuclei; doublets were then removed using Scrublet[57]; finally, nuclei and genes were filtered using the same quality-control criteria as the Visium data. The filtered gene-expression matrix was normalized, and batch effects were removed using bbknn. Unsupervised clustering was performed with the Leiden algorithm (resolution = 1.2), and two-dimensional projection was generated with UMAP (min_dist = 0.4). Initial annotation was carried out using CellAssign, after which final cell-type annotations were determined manually based on the expression of marker genes.

Identification of DEGs between cell types in Visium spatial transcriptomics data

Following quality control, normalization, and cell-type annotation, DEGs for each cell type were identified using the scanpy.tl.rank_genes_groups function within the Scanpy framework. A "one-vs-rest" comparison strategy was employed using the normalized expression matrix (use_raw=False). Statistical significance was evaluated using the *t*-test with Benjamini-Hochberg correction. Genes were defined as significant DEGs if they met the following criteria: a log2foldchange > 2, an adjusted *p*-value < 0.05, and a fraction of cells expressing the gene (pts) > 0.8 in the target cell type.

Identification of ASEs and imprinted genes using Visium Spatial and snRNA-seq

transcriptomic data

The workflow for identifying ASE and imprinted genes primarily consists of three steps: high-quality SNP calling, parent-of-origin phasing based on N-masked genome, and calculation of parental expression matrices followed by imprinted gene identification.

1) SNP identification between B73 and Mo17

To obtain high-confidence SNPs between B73 (Zm-B73-REFERENCE-NAM-5.0) and Mo17 (Zm-Mo17-REFERENCE-CAU-2.0) [32], we performed chromosome-to-chromosome alignment using nucmer[58] and extracted SNPs using the show-snps utility. In parallel, high-coverage HiFi reads of Mo17 were downloaded, aligned to B73 genome using minimap2[59] and SNPs were called following the standard GATK pipeline[60]. SNPs identified simultaneously by nucmer and GATK-based variant calling were considered high-confidence SNPs and used for downstream phasing.

2) Parent-of-origin phasing based on N-masked genome

The B73 v5 genome was first masked at all high-quality SNP positions identified in the steps above, by replacing the corresponding nucleotides with "N". The read2 of 10x Genomics Visium and DNBelab C4 snRNA-seq pair-end data were aligned to this N-masked genome using STAR[61] with parameters --alignEndsType EndToEnd --outSAMattributes NH HI NM MD --outSAMtype BAM Unsorted. The bam file was sorted by sambamba[62], then filtered to retained the unique alignments. Then, the bam file was split into two separate bam files using SNPsplit[63] (--single_end --conflicting) and high-quality SNPs between B73 and Mo17. The resulting *genome1.bam was

regarded as the B73-origin bam file, and *genome2.bam was regarded as the Mo17-origin bam file. Read names were extracted from the BAM files using seqtk and used to extract the phased subsets from the Visium and snRNA-seq FASTQ data.

3) The calculation of parental expression matrix and identification of imprinted genes for each cell type

The phased FASTQ files were processed using the standard Space Ranger pipeline for the Visium data, and the dnbc4tools pipeline for the snRNA-seq data, to generate gene expression matrices. ASE analysis was subsequently performed at the cell-type level. Specifically, for each annotated cell type, genes with at least 20 allele-assigned reads (combining B73- and Mo17-derived reads) across all spots were analyzed. The B73BiasRatio for each gene was calculated using distinct formulas for the embryo cell types and endosperm cell types to account for their genomic dosage differences. Specifically, for the endosperm cell types,

$$B73BiasRatio = \begin{cases} \dfrac{B73counts}{B73counts + 2 \times Mo17counts}, & if\,\mathrm{BM} \\ \dfrac{B73counts}{2 \times B73counts + Mo17counts}, & if\,\mathrm{MB} \end{cases}$$

whereas for the embryo cell types,

$$B73BiasRatio = \frac{B73counts}{B73counts + Mo17counts}$$

genes with B73BiasRatio $\geqslant$ 0.9 were defined as B73-biased ASE genes, while those with B73BiasRatio $\leqslant$ 0.1 were defined as Mo17-biased ASE genes. For genes within the same cell type, those meeting the following criteria were defined as MEG:

exhibiting B73-ASE in the BM cross and Mo17-ASE in the MB cross, with parental read counts significantly deviating from the expected 2:1 ratio in a chi-square test. Conversely, genes showing Mo17-ASE in the BM cross and B73-ASE in the MB cross, along with a statistically significant departure from the 2:1 ratio in the chi-square test, were defined as PEG. For the snRNA-seq data, a gene was retained as a MEG or PEG only if it met the above criteria in both replicates.

WGBS data processing and read phasing

Raw FASTQ files were quality-filtered using fastp[64] to generate clean reads, and SNPsplit[63] (--paired -bisulfit --conflicting) was used for haplotype phasing. The haplotype-assigned clean reads were subsequently aligned to the B73 reference genome (Zm-B73-REFERENCE-NAM-5.0) using Bismark[65] (0.24.2). PCR duplicates were removed with deduplicate_bismark, and cytosine methylation information was extracted using bismark_methylation_extractor with parameters --comprehensive --cytosine_report --CX_context --split_by_chromosome to obtain the number of methylated and unmethylated reads at each site.

GO enrichment analysis

Gene Ontology (GO) annotations for B73_RefGen_v5 were downloaded from MaizeGDB (B73_GO.out). Only entries with ARGOT_PPV $\geqslant$ 0.5 were retained. GO enrichment analysis was performed using the enrichr module of GSEApy[66]. GO terms

with *P*-value ≤ 0.05 were considered significantly enriched and were ranked in ascending order of *P*-values for visualization.

Identification and quantification of lncRNAs

To enable accurate quantification of lncRNAs in spatial transcriptomic data, we performed genome-guided lncRNA annotation using full-length transcriptome data from embryos and endosperms of both BM and MB samples. Specifically, the following pipeline was applied to generate high confidence lncRNA annotations based on FLNC reads: (1) Cleaned BAM files from the four tissue samples were merged; (2) StringTie2[67] was used with the -L and -G options for reference-guided transcript assembly to produce a GTF file; (3) Candidate lncRNAs were filtered based on the following criteria: transcript length > 200 bp; coding sequence < 100 amino acids; no overlap with known annotations; and correct strand prediction; (4) CPC2[42] was used to evaluate coding potential, and only transcripts lacking coding potential were retained as high-confidence lncRNAs. The newly annotated lncRNAs were merged with the B73_RefGen_v5 GTF annotation. Space Ranger and dnbc4tools were re-run using the updated annotation to generate transcriptomic count matrices that include lncRNAs. lncRNAs were subsequently analyzed for ASE and imprinting using the same criteria applied to coding genes.

Tau calculation and gene classification

Tau values for both coding genes and lncRNAs were calculated using the tspex package to evaluate cell-type specificity. Based on these values, genes were classified into three categories: cell-type specific (Tau between 0 and 0.3), intermediate (Tau between 0.3 and 0.7), and ubiquitous (Tau between 0.7 and 1).

Pseudotime analysis

Pseudotime analysis was performed separately on the snRNA-seq data from the BM and MB crosses using Monocle2[43]. Specifically, the expression matrices were processed using the same normalization and dispersion estimation procedures as described above. Highly variable genes were identified based on mean expression (≥0.1) and empirical dispersion (≥1.0) and were used as ordering genes for trajectory inference. Dimensionality reduction was performed using the DDRTree algorithm implemented in Monocle 2, followed by cell ordering along the inferred trajectories with the orderCells function to assign pseudotime values. For the allele-specific expression matrices obtained from parental phasing of each sample, the same analysis pipeline was applied separately to generate maternal and paternal pseudotime trajectories. The parental pseudotime difference was then defined as the difference between the maternal and paternal pseudotime values for each nucleus.

Supporting information

Supp.Table.1 Basic statistics of spatial and single-nucleus transcriptome sequencing

of hybrid maize

Supp.Table.2 Cell-type specific DEGs identified from spatial transcriptomics

Supp.Table.3 PMR and effective read counts of ASE genes across different cell types

Supp.Table.4 PMR and effective read counts of imprinted genes across different cell types

Supp.Table.5 Imprinted genes identified by single-nucleus transcriptomics

Declaration of interests

The authors declare no competing interests.

Code availability

The code used in this study is available at Github (https://github.com/litao1503/Hybrid_Maize_Spatial_Transcriptomics)

Data availability

The 10x Visium spatial transcriptomics, DNBelab C4 snRNA transcriptomics and WGBS datasets generated in this study have been deposited in NCBI under BioProject accession PRJNA1290227.

Acknowledgements

We acknowledge the High-Performance Computing Public Platform (Shenzhen Campus) of Sun Yat-sen University for computational support. We also want to thank Prof. Zhang, Mei in the Institute of Botany, The Chinese Academy of Science (CAS), and Dr. Xiao, Wenxin in the CAS Center for Excellence in Molecular Plant Science for their suggestion and assistance during project design and data analysis. This work was supported by grants from the Shenzhen Outstanding Youth Science Fund project (RCYX20231211090225038), the Shenzhen Science and Technology Program (KCXFZ20240903093007010), the National Natural Science Foundation of China (32372115), and the Open Project of State Key Laboratory of Maize Bio-breeding (SKLMB2025K01) to J.S.

Author contributions

J.S. conceived and supervised the overall research project. T.L. performed most data analyses and prepared the figures. M.Z., T.L., and Y.J. collected the plant materials. Y.J., K.Z., S.J., X.Y., H.L., J.W., and X.D. contributed to the data analysis and interpretation. T.L. and J.S. wrote the manuscript, and J.S. finalized the manuscript. All authors read and approved the manuscript.

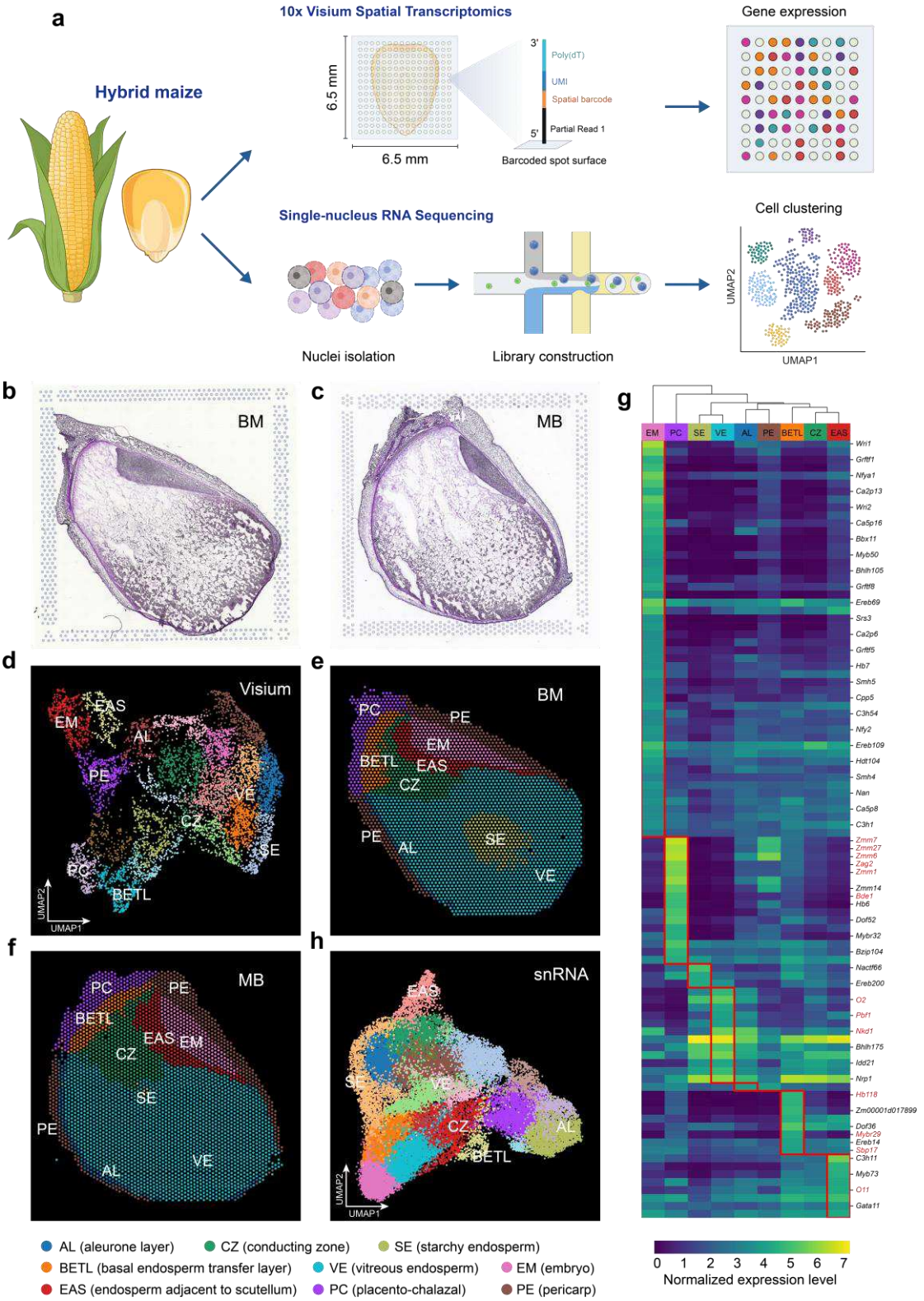
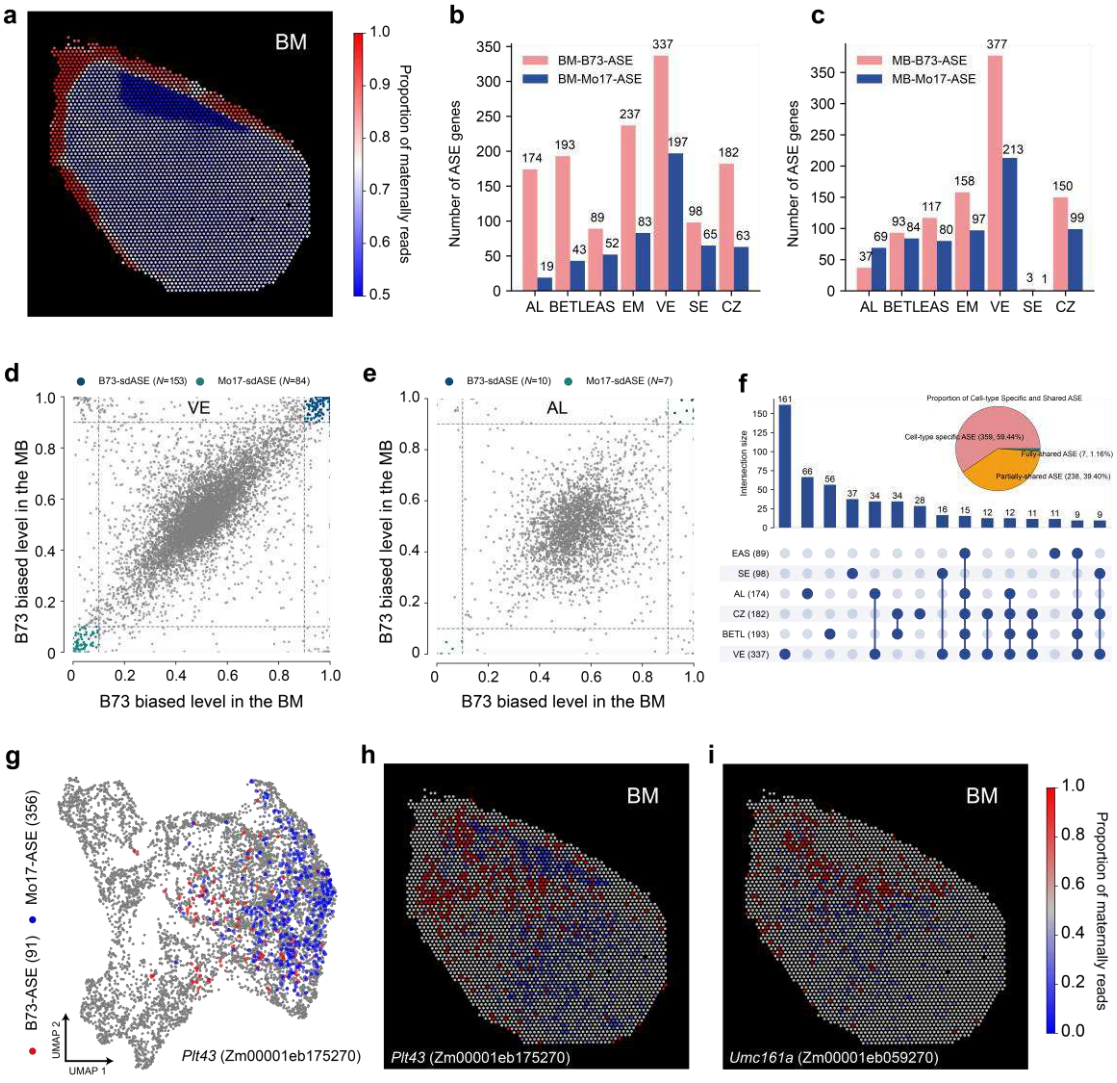


Fig. 1 Spatial and single-nucleus transcriptomic profiling of maize reciprocal hybrid kernels.

a, Schematic of the experimental design. **b, c** The anatomical structure under

hematoxylin and eosin staining of BM (b) and MB (c) kernels at their maximal plane in 14 DAP. **d** UMAP projection of all spatial transcriptomic spots integrating the BM and MB crosses, colored by unsupervised cluster. **e, f** The annotation of spatial compartment (labeled in the bottom) for BM (e) and MB (f) samples. **g** The heatmap of transcription factor genes (N = 98) that differentially expressed across nine cell types. **h** UMAP projection of snRNA-seq data with each unsupervised cluster labeled by different colors.



678

679 **Fig. 2 Identification and characterization of ASE genes across different cell types.**

680 **a** Spatial plot showing the PMR of each spot in the BM sample. **b, c** Bar plot showing

681 the number of B73-ASE and Mo17-ASE genes identified in each cell type of BM (b)

682 and MB (c). **d, e** Scatter plot showing the distribution of sdASE genes in VE (d) and

683 AL (e). Blue and green dots represent B73-sdASE and Mo17-sdASE genes, respectively.

684 **f** UpSet plot showing the overlaps of B73-ASE genes across different cell types in BM.

685 Pie chart classifies ASE genes into three categories: cell-type specific ASE (identified

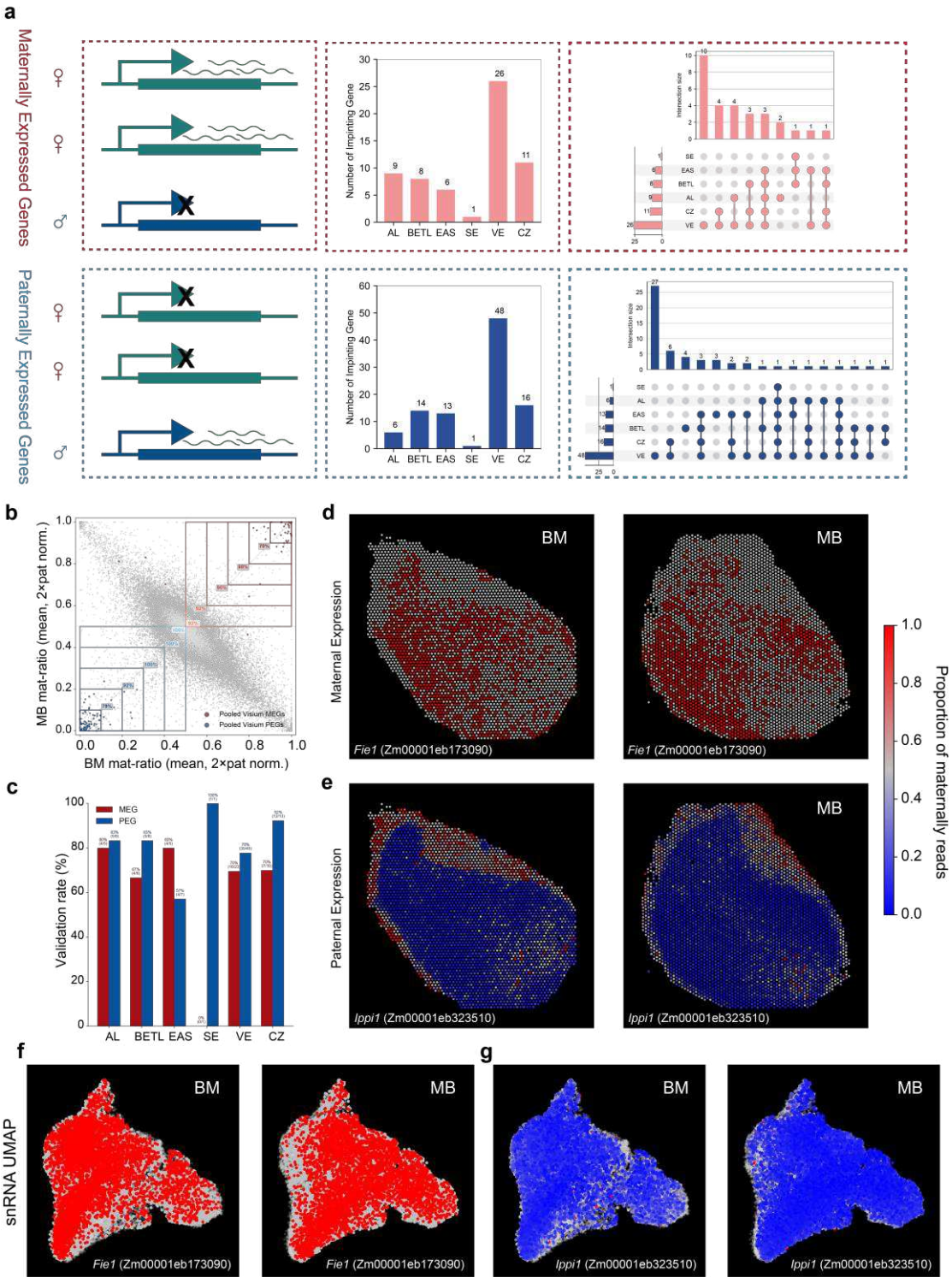
686 in only one cell type), partially shared ASE (shared in more than one but not all cell

687 types), and fully shared ASE (shared across all cell types). **g, h** UMAP (g) and spatial

688 PMR map (h) of *Plt43* (Zm00001eb175270) gene in BM. **i** Spatial PMR map of

689 *Umc161a* (Zm00001eb059270) gene in BM.

690



691

692 **Fig. 3 Cellular and spatial specificity of genomic imprinting in maize endosperm.**

693 **a** Schematic representation of MEG and PEG in maize endosperm (left). Bar plot

694 showing the number of MEGs and PEGs identified in each cell type (middle) in Visium

695 transcriptomic data. UpSet plot showing the overlap of MEGs and PEGs across cell

types (right). **b** Scatter plot showing the maternal ratio of analyzable Visium PEGs and MEGs in the snRNA-seq data. The box and proportions shown inside represent the overall validation rate of PEGs and MEGs under different criteria for calling imprinted genes. **c** Bar plot showing the proportion of analyzable Visium PEGs and MEGs in six endosperm cell types that were validated by the snRNA-seq data. **d, e** Spatial PMR plots showing imprinting status of the MEG *Fie1* (d) and PEG *Ippi1* (e) in BM (left) and MB (right). **f, g** UMAP PMR plots showing imprinting status of the MEG *Fie1* (f) and PEG *Ippi1* (g) in BM (left) and MB (right).

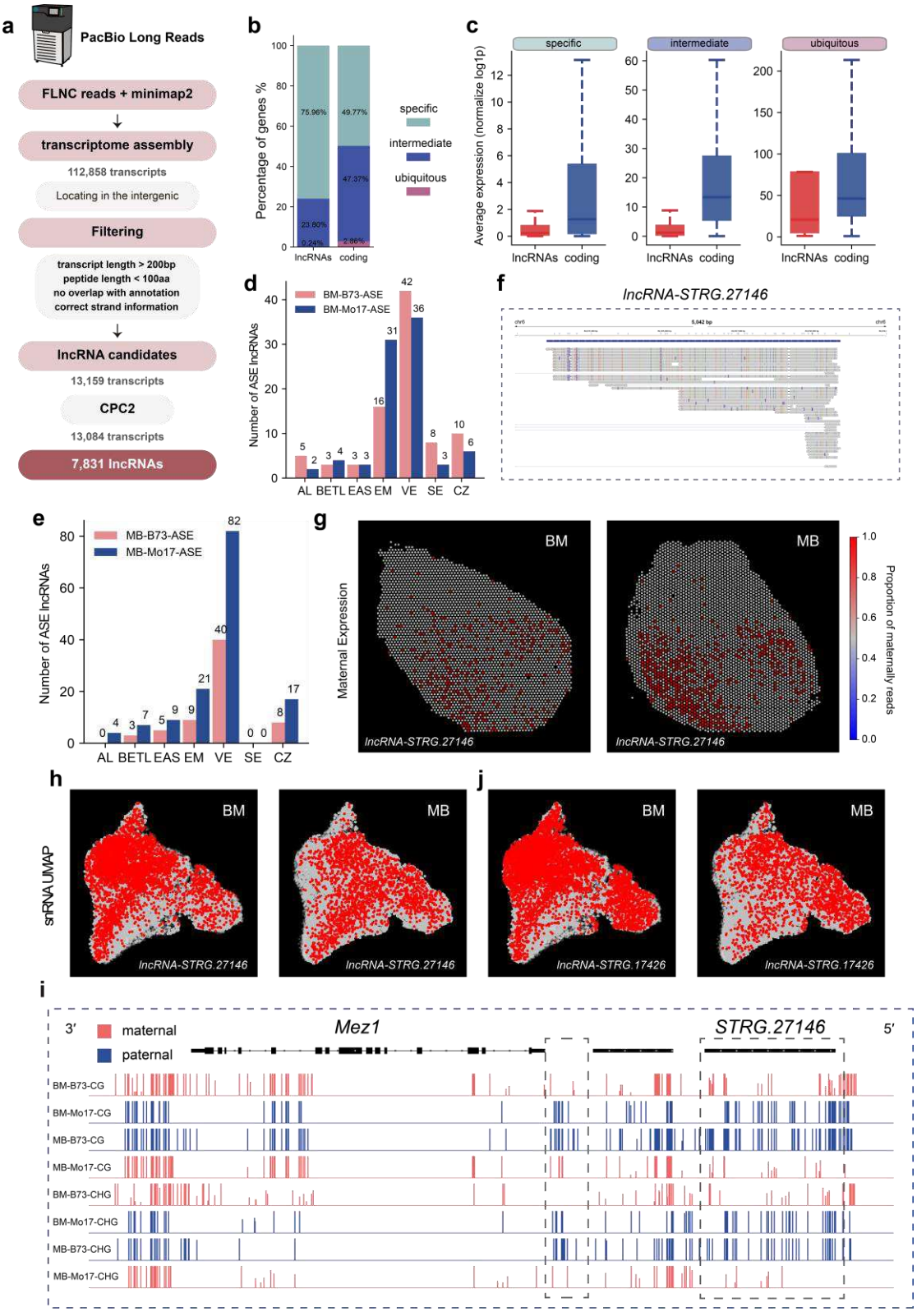


Fig. 4 Spatial and cellular characterization of imprinted lncRNAs in maize endosperm.

a Workflow for lncRNA identification using the long-read transcriptome data. **b, c** The

proportion (b) and expression levels (c) of lncRNAs and protein-coding genes across three categories according to cell-type specificity: tissue-specific, intermediate, and ubiquitous. **d, e** Bar plot showing the number of ASE lncRNAs in each cell type of BM (d) and MB (e). **f** IGV visualization of long cDNA reads aligned in the genomic region of *STRG.27146*. **g** Spatial PMR plot showing imprinting status of the maternally expressed *STRG.27146* in BM (left) and MB (right). **h** PMR distribution of snRNA-seq UMAP showing imprinting status of the maternally expressed *STRG.27146* in BM (left) and MB (right). **i** IGV plot showing allele-resolved CG and CHG methylation levels of the *Mez1* and *STRG.27146* loci in reciprocal hybrid endosperms. Red and blue tracks indicate maternal and paternal alleles, respectively. **j** PMR distribution of snRNA-seq UMAP showing imprinting status of the maternally expressed *STRG.17426* in BM (left) and MB (right).

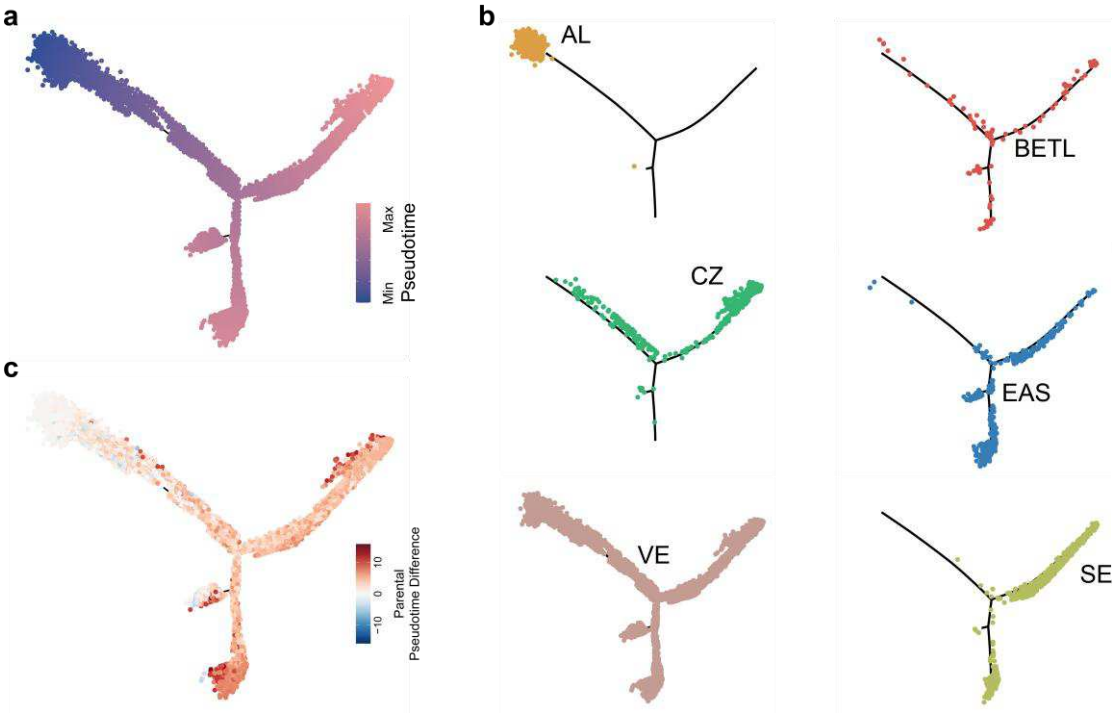


Fig. 5 Haplotype-resolved single-nucleus pseudotime analysis and developmental trajectory of maize endosperm.

a Single-nucleus pseudotime analysis of BM endosperm nuclei using Monocle2. **b** Distribution of the six endosperm cell types (AL, BETL, CZ, EAS, VE and SE) along the snRNA-seq pseudotime trajectory in BM. **c** Parental pseudotime difference projected onto the BM snRNA-seq developmental trajectory. For each nucleus, the parental pseudotime difference was calculated as the maternal-allele-derived pseudotime minus the paternal-allele-derived pseudotime.

References

1 Doll, N. M., Depège-Fargeix, N., Rogowsky, P. M. & Widiez, T. Signaling in early maize kernel development. *Molecular Plant* 10, 375-388 (2017).

2 Dai, D., Ma, Z. & Song, R. Maize endosperm development. *Journal of Integrative Plant Biology* 63, 613-627 (2021).

3 Dai, D., Ma, Z. & Song, R. Maize kernel development. *Molecular Breeding* 41, 1-33 (2021).

4 McClintock, B. The origin and behavior of mutable loci in maize. *Proceedings of the National Academy of Sciences* 36, 344-355 (1950).

5 Kermicle, J. Dependence of the R-mottled aleurone phenotype in maize on mode of sexual transmission. *Genetics* 66, 69 (1970).

6 Gehring, M. Genomic imprinting: insights from plants. *Annual review of genetics* 47, 187-208 (2013).

7 Gutieérrez-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. *The Plant Cell* 16, 1288-1301 (2004).

8 Costa, L. M. *et al.* Maternal control of nutrient allocation in plant seeds by genomic imprinting. *Current Biology* 22, 160-165 (2012).

9 Luo, M., Platten, D., Chaudhury, A., Peacock, W. & Dennis, E. S. Expression, imprinting, and evolution of rice homologs of the polycomb group genes. *Molecular plant* 2, 711-723 (2009).

10 Huang, X. *et al.* Imprinted gene OsFIE1 modulates rice seed development by influencing nutrient metabolism and modifying genome H3K27me3. *The Plant Journal* 87, 305-317 (2016).

11 Cheng, X. *et al.* Maternally expressed FERTILIZATION-INDEPENDENT ENDOSPERM1 regulates seed dormancy and aleurone development in rice. *The Plant Cell* 37, koae304 (2025).

12 Danilevskaya, O. N. *et al.* Duplicated fie genes in maize: expression pattern and imprinting suggest distinct functions. *The Plant Cell* 15, 425-438 (2003).

13 Hermon, P., Srilunchang, K.-o., Zou, J., Dresselhaus, T. & Danilevskaya, O. N. Activation of the imprinted Polycomb Group Fie1 gene in maize endosperm requires demethylation of the maternal allele. *Plant molecular biology* 64, 387-395 (2007).

14 Yang, J. *et al.* The role of imprinted gene ZmFIE1 during maize kernel development. *The Crop Journal* 13, 395-405 (2025).

15 Zhang, M. *et al.* Extensive, clustered parental imprinting of protein-coding and noncoding RNAs in developing maize endosperm. *Proceedings of the National Academy of Sciences* 108, 20042-20047 (2011).

16 Waters, A. J. *et al.* Comprehensive analysis of imprinted genes in maize reveals allelic variation for imprinting and limited conservation with other species. *Proceedings of the National Academy of Sciences* 110, 19639-19644

(2013).
17 Zhang, M. *et al.* Genome-wide high resolution parental-specific DNA and histone methylation maps uncover patterns of imprinting regulation in maize. *Genome research* 24, 167-176 (2014).
18 Dong, X. *et al.* Dynamic and antagonistic allele-specific epigenetic modifications controlling the expression of imprinted genes in maize endosperm. *Molecular plant* 10, 442-455 (2017).
19 Dong, X. *et al.* Parent-of-origin-dependent nucleosome organization correlates with genomic imprinting in maize. *Genome Research* 28, 1020-1028 (2018).
20 Dai, D., Ma, Z. & Song, R. Maize kernel development. *Molecular Breeding* 41, 2 (2021).
21 Fu, Y. *et al.* Spatial transcriptomics uncover sucrose post-phloem transport during maize kernel development. *Nature communications* 14, 7191 (2023).
22 Rodrigues, J. A. & Zilberman, D. Evolution and function of genomic imprinting in plants. *Genes & development* 29, 2517-2531 (2015).
23 Yuan, Y. *et al.* Decoding the gene regulatory network of endosperm differentiation in maize. *Nature communications* 15, 34 (2024).
24 Cohen, M. *et al.* Lung single-cell signaling interaction map reveals basophil role in macrophage imprinting. *Cell* 175, 1031-1044. e1018 (2018).
25 Mazzurana, L. *et al.* Tissue-specific transcriptional imprinting and heterogeneity in human innate lymphoid cells revealed by full-length single-cell RNA-sequencing. *Cell research* 31, 554-568 (2021).
26 Santoni, F. A. *et al.* Detection of imprinted genes by single-cell allele-specific gene expression. *The American Journal of Human Genetics* 100, 444-453 (2017).
27 Stewart, O. J. *et al.* Monoallelic expression can govern penetrance of inborn errors of immunity. *Nature*, 1-12 (2025).
28 Laukoter, S. *et al.* Cell-type specificity of genomic imprinting in cerebral cortex. *Neuron* 107, 1160-1179. e1169 (2020).
29 Wu, J.-J. *et al.* Cell-cell communication-mediated cell-type-specific parent-of-origin effects in mammals. *Nature Communications* 16, 1-17 (2025).
30 Picard, C. L., Povilus, R. A., Williams, B. P. & Gehring, M. Transcriptional and imprinting complexity in Arabidopsis seeds at single-nucleus resolution. *Nature plants* 7, 730-738 (2021).
31 Jiao, Y. *et al.* Improved maize reference genome with single-molecule technologies. *Nature* 546, 524-527 (2017).
32 Chen, J. *et al.* A complete telomere-to-telomere assembly of the maize genome. *Nature genetics* 55, 1221-1231 (2023).
33 Polański, K. *et al.* BBKNN: fast batch alignment of single cell transcriptomes. *Bioinformatics* 36, 964-965 (2020).
34 Tian, R. *et al.* Regulation of maize kernel development via divergent

activation of α-Zein genes by transcription factors O11, O2, and PBF1. *Journal of Genetics and Genomics* (2025).

35 Sun, Y. *et al.* Progressive meristem and single-cell transcriptomes reveal the regulatory mechanisms underlying maize inflorescence development and sex differentiation. *Molecular Plant* 17, 1019-1037 (2024).

36 Raissig, M. T., Baroux, C. & Grossniklaus, U. Regulation and flexibility of genomic imprinting during seed development. *The Plant Cell* 23, 16-26 (2011).

37 Dai, D. *et al.* Paternal imprinting of dosage-effect defective1 contributes to seed weight xenia in maize. *Nature Communications* 13, 5366 (2022).

38 Meng, D. *et al.* Sequential gene activation and gene imprinting during early embryo development in maize. *The Plant Journal* 93, 445-459 (2018).

39 Batista, R. A. & Köhler, C. Genomic imprinting in plants—revisiting existing models. *Genes & development* 34, 24-36 (2020).

40 Xu, G. & Zhang, X. Mechanisms controlling seed size by early endosperm development. *Seed Biology* 2, 1-11 (2023).

41 Tucci, V. *et al.* Genomic imprinting and physiological processes in mammals. *Cell* 176, 952-965 (2019).

42 Kang, Y.-J. *et al.* CPC2: a fast and accurate coding potential calculator based on sequence intrinsic features. *Nucleic acids research* 45, W12-W16 (2017).

43 Qiu, X. *et al.* Reversed graph embedding resolves complex single-cell trajectories. *Nature methods* 14, 979-982 (2017).

44 Wolff, P. *et al.* High-resolution analysis of parent-of-origin allelic expression in the Arabidopsis endosperm. *PLoS genetics* 7, e1002126 (2011).

45 Xin, M. *et al.* Dynamic expression of imprinted genes associates with maternally controlled nutrient allocation during maize endosperm development. *The Plant Cell* 25, 3212-3227 (2013).

46 Van Ekelenburg, Y. S. *et al.* Spatial and temporal regulation of parent-of-origin allelic expression in the endosperm. *Plant physiology* 191, 986-1001 (2023).

47 Zhou, H., Deng, X. W. & He, H. Gene expression variations and allele-specific expression of two rice and their hybrid in caryopses at single-nucleus resolution. *Frontiers in Plant Science* 14, 1171474 (2023).

48 Longo, S. K., Guo, M. G., Ji, A. L. & Khavari, P. A. Integrating single-cell and spatial transcriptomics to elucidate intercellular tissue dynamics. *Nature Reviews Genetics* 22, 627-644 (2021).

49 Luo, C. *et al.* Single-cell methylomes identify neuronal subtypes and regulatory elements in mammalian cortex. *Science* 357, 600-604 (2017).

50 Marand, A. P., Chen, Z., Gallavotti, A. & Schmitz, R. J. A cis-regulatory atlas in maize at single-cell resolution. *Cell* 184, 3041-3055. e3021 (2021).

51 Deng, Y. *et al.* Spatial profiling of chromatin accessibility in mouse and

human tissues. *Nature* 609, 375-383 (2022).
52 Marand, A. P. *et al.* The genetic architecture of cell type–specific cis regulation in maize. *Science* 388, eads6601 (2025).
53 Singh, M. *et al.* High-throughput targeted long-read single cell sequencing reveals the clonal and transcriptional landscape of lymphocytes. *Nature communications* 10, 3120 (2019).
54 Wolf, F. A., Angerer, P. & Theis, F. J. SCANPY: large-scale single-cell gene expression data analysis. *Genome biology* 19, 1-5 (2018).
55 Gayoso, A. *et al.* A Python library for probabilistic analysis of single-cell omics data. *Nature biotechnology* 40, 163-166 (2022).
56 Zhang, A. W. *et al.* Probabilistic cell-type assignment of single-cell RNA-seq for tumor microenvironment profiling. *Nature methods* 16, 1007-1015 (2019).
57 Wolock, S. L., Lopez, R. & Klein, A. M. Scrublet: computational identification of cell doublets in single-cell transcriptomic data. *Cell systems* 8, 281-291. e289 (2019).
58 Marçais, G. *et al.* MUMmer4: A fast and versatile genome alignment system. *PLoS computational biology* 14, e1005944 (2018).
59 Li, H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* 34, 3094-3100 (2018).
60 Poplin, R. *et al.* Scaling accurate genetic variant discovery to tens of thousands of samples. *BioRxiv*, 201178 (2017).
61 Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 29, 15-21 (2013).
62 Tarasov, A., Vilella, A. J., Cuppen, E., Nijman, I. J. & Prins, P. Sambamba: fast processing of NGS alignment formats. *Bioinformatics* 31, 2032-2034 (2015).
63 Krueger, F. & Andrews, S. R. SNPsplit: Allele-specific splitting of alignments between genomes with known SNP genotypes. *F1000Research* 5, 1479 (2016).
64 Chen, S., Zhou, Y., Chen, Y. & Gu, J. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* 34, i884-i890 (2018).
65 Krueger, F. & Andrews, S. R. Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. *bioinformatics* 27, 1571-1572 (2011).
66 Fang, Z., Liu, X. & Peltz, G. GSEApy: a comprehensive package for performing gene set enrichment analysis in Python. *Bioinformatics* 39, btac757 (2023).
67 Kovaka, S. *et al.* Transcriptome assembly from long-read RNA-seq alignments with StringTie2. *Genome biology* 20, 1-13 (2019).

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- ImprintingSuppTab0616.xlsx
- ImprintingSuppFig0616.docx