

RESEARCH

Open Access



Single-cell RNA-seq analysis reveals ploidy-dependent and cell-specific transcriptome changes in *Arabidopsis* female gametophytes

Qingxin Song^{1,2†}, Atsumi Ando^{1†}, Ning Jiang³, Yoko Ikeda⁴ and Z. Jeffrey Chen^{1*} 

* Correspondence: zjchen@austin.utexas.edu

[†]Qingxin Song and Atsumi Ando contributed equally to this work.

¹Department of Molecular Biosciences, The University of Texas at Austin, 1 University Station A5000, Austin, TX 78712, USA
Full list of author information is available at the end of the article

Abstract

Background: Polyploidy provides new genetic material that facilitates evolutionary novelty, species adaptation, and crop domestication. Polyploidy often leads to an increase in cell or organism size, which may affect transcript abundance or transcriptome size, but the relationship between polyploidy and transcriptome changes remains poorly understood. Plant cells often undergo endoreduplication, confounding the polyploid effect.

Results: To mitigate these effects, we select female gametic cells that are developmentally stable and void of endoreduplication. Using single-cell RNA sequencing (scRNA-seq) in *Arabidopsis thaliana* tetraploid lines and isogenic diploids, we show that transcriptome abundance doubles in the egg cell and increases approximately 1.6-fold in the central cell, consistent with cell size changes. In the central cell of tetraploid plants, *DEMETER* (*DME*) is upregulated, which can activate *PRC2* family members *FIS2* and *MEA*, and may suppress the expression of other genes. Upregulation of cell size regulators in tetraploids, including *TOR* and *OSR2*, may increase the size of reproductive cells. In diploids, the order of transcriptome abundance is central cell, synergid cell, and egg cell, consistent with their cell size variation. Remarkably, we uncover new sets of female gametophytic cell-specific transcripts with predicted biological roles; the most abundant transcripts encode families of cysteine-rich peptides, implying roles in cell-cell recognition during double fertilization.

Conclusions: Transcriptome in single cells doubles in tetraploid plants compared to diploid, while the degree of change and relationship to the cell size depends on cell types. These scRNA-seq resources are free of cross-contamination and are uniquely valuable for advancing plant hybridization, reproductive biology, and polyploid genomics.

Keywords: Polyploidy, Single-cell RNA-seq, Transcriptome, Gametogenesis, Reproduction



© The Author(s). 2020 **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 licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence 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 licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

Introduction

Polyploidy or whole-genome duplication (WGD) is a widespread phenomenon that has dominated the genome evolution of many animals and all flowering plants [1–6]. Moreover, polyploid cells can form through endoreduplication during development in otherwise diploid organisms including humans [7]. The common occurrence of polyploids suggests an advantage of having additional genetic materials for evolution, adaptation, and domestication [1, 3, 5, 8, 9]. For example, yeast polyploids can obtain rapid adaptation through higher rates of beneficial mutations [10]. *Arabidopsis* autotetraploids have enhanced salinity tolerance, which is associated with elevated potassium and reduced sodium levels [11]. In *Arabidopsis* allotetraploids and *Arabidopsis thaliana* hybrids, epigenetic changes induce altered circadian rhythms, which increases photosynthesis and starch metabolism [12] and gates the timing of stress responses [13] and ethylene production [14], leading to increased growth traits such as biomass heterosis [15].

Polyploidy often leads to cell size increase as observed in yeast and *Arabidopsis* [16, 17]. However, results from gene expression studies on yeast and plant autopolyploids are inconsistent [17–20]. In yeast, ploidy variation alters a dozen of genes that regulate cell cycles and cell surface [17], while the number of genes whose expression is altered by tetraploidy varies from nine to several hundreds among different *A. thaliana* ecotypes [19, 20]. In *Glycine* species using genomic DNA normalization, the tetraploid has a 1.4-fold transcriptome abundance relative to its diploid and exhibits dosage effects on the majority of expressed genes [18, 21]. A recent study using sorted endoreduplicated nuclei in tomato fruits of diploid plants has shown a genome-wide proportional shift of gene expression depending on ploidy levels [22]. These different results may suggest that polyploid effects on gene expression vary from one genotype to another or one organism to another. Alternatively, technological limitations such as RNA-seq and microarray assays often examine relative gene expression levels and may not measure the absolute transcript abundance per gene per cell [21, 23, 24].

Single-cell RNA sequencing (scRNA-seq) analysis provides an effective alternative to study the polyploid effects on absolute levels of gene expression changes because it allows quantifying absolute transcript numbers of individual genes per cell for all genes in the genome [25, 26]. The scRNA-seq approach has been extensively used to map transcriptome dynamics from human embryos [27] to tumor evolution [28]. However, the progress on plant single-cell genomics is limited [29, 30], and transcriptome changes in polyploid plants at the single-cell level are unknown [31]. In this study, we have employed scRNA-seq technique to map absolute transcript dynamics in female gametophytic cells of *A. thaliana* diploid and isogenic autotetraploid plants, whose ploidy levels have been validated in other studies [16, 32]. The scRNA-seq results have shown ploidy-dependent and cell type-specific effects on transcriptome changes and provided unique gene expression features and valuable resources that are free of cross-contamination in the egg, central, and synergid cells during female gametophytic development.

Results

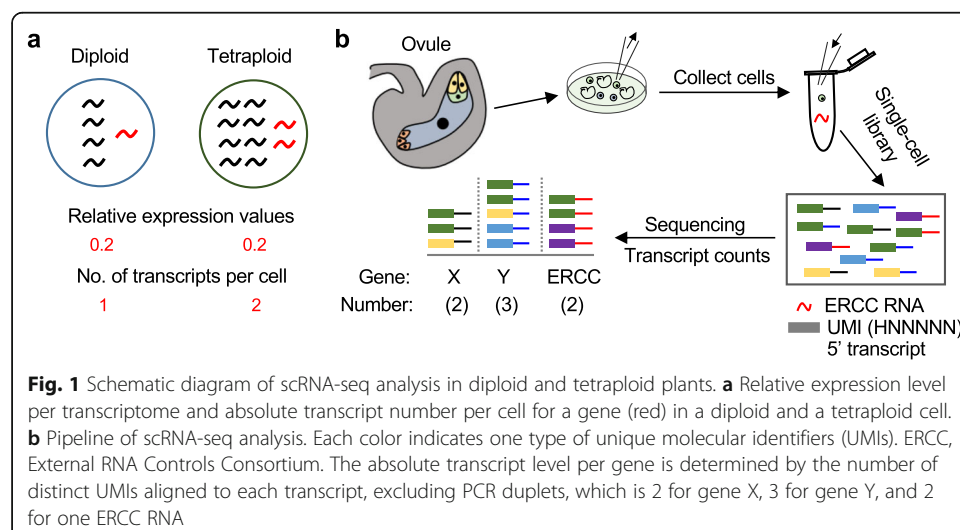
Experimental validation for single-cell analysis in female gametic cells

A tetraploid cell has twice the amount of DNA relative to a diploid cell, but transcriptome studies have found a small number of genes showing expression changes between tetraploids and diploids in *Arabidopsis* [19, 20]. This is likely caused by measuring

relative gene expression levels that cannot accurately measure transcriptome abundance between polyploid and diploid cells [21]. For example, if the total RNA amount doubles in the tetraploid cell relative to the diploid, the absolute number of gene transcripts would exhibit averagely twofold increase in the tetraploid relative to the diploid (Fig. 1a), whereas the relative expression level of the gene per transcriptome would be identical between diploid and tetraploid cells. In addition, some normalization methods such as spike-in RNA or genomic DNA [18, 23, 33] measure the expression level of each gene to the control (such as spike-in RNA or genomic DNA/ploidy), but not the absolute transcript numbers per gene per cell (Additional file 1: Fig. S1).

To solve this problem, we employed the scRNA-seq approach to quantify the genome-wide changes in absolute transcript levels per gene per cell in diploid and tetraploid plants (Fig. 1b). The choice of cell types is critical for the scRNA analysis. Many plant cell types often undergo endoreduplication [22], which can confound the polyploid (WGD) effect. After evaluating suitable cell types, we selected reproductive (female gametophytic) cells for the study to minimize confounding variables, and these cells are relatively uniform in a given stage. Flowering plants have three distinct female gametic cells, namely, egg, central, and synergid cells. Each egg or synergid cell is a haploid, whereas the central cell is a diploid (2 copies of the maternal genome). We generated two cell-specific lines that each expresses green fluorescent protein (GFP) in the diploid or isogenic tetraploid *A. thaliana* (Col-0) using the same construct as previously reported [34, 35]. Specifically, pDD45:nGFP and pSUP16:nGFP are expressed in the egg and central cell nuclei, respectively [34]. The synergid-cell *GFP* line was available only in the diploid plants and used for diploid comparison but excluded from the diploid-tetraploid analysis.

Using the nGFP marker, we manually isolated each egg cell from an ovule of flowers (stage 12) under an inverted dissecting microscope (the “Methods” section). We named the egg cell in the diploid (ECd) and tetraploid (ECt) plants (Fig. 2a). Similarly, the central cells (CCs) from the diploid and tetraploid plants are designated CCd and CCt, respectively (Fig. 2b). The egg and central cells in the tetraploid plants were ~ 1.6-fold (by volume) larger than the corresponding cells in their corresponding diploids (Fig. 2c,

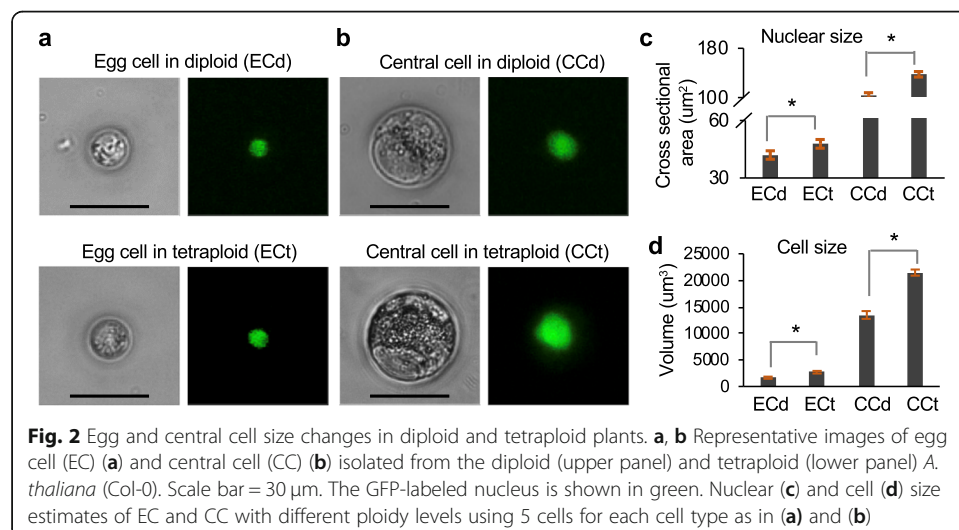


d), which is consistent with the cell size increase in yeast polyploids [17] and in stomatal cells of *A. thaliana* ploidy series [16]. Note that CCs are noticeably larger than the ECs in the same ovule probably because each CC is a diploid (two copies of the maternal genome) in a diploid plant or tetraploid in a tetraploid plant, while each EC is a haploid and diploid in the diploid and tetraploid plants, respectively.

For egg cell and central cell in the diploid and tetraploid plants, scRNA-seq libraries were constructed using eight cells (eight biological replicates) for each cell type as previously described [26, 36] (the “Methods” section). In addition, synergid cells in the diploid plants with four biological replicates were used for scRNA-seq library construction and analysis. An equal amount of External RNA Controls Consortium (ERCC) RNA molecules was added to each cell (Fig. 1b). During reverse transcription, each cDNA molecule was ligated with a unique molecular identifier (UMI) containing 6-bp random sequence (HNNNNN) at the 5' end, which could be used to correct PCR-induced amplification bias, while the ligation efficiency was estimated using the external RNA controls. The raw transcript number per gene was calculated as the number of distinct UMIs aligned to the gene model, excluding duplicate counts.

Another quality control showed that the majority of sequencing reads were aligned to the transcription start sites of both endogenous genes and ERCC RNAs, indicating good preservation of mRNA integrity during RNA manipulation (Additional file 1: Fig. S2). For the expression of ERCC RNAs, the dependence of the squared coefficients of variation (CV^2) on the molecule counts fits the Poisson distribution, which is consistent with previous single-cell RNA-seq studies [25, 37] (Additional file 1: Fig. S3). Moreover, patterns of the female gamete specifically expressed genes (6 in egg cell and 12 in central cell) from the published datasets [38–41] were reproduced in our scRNA-seq datasets (Additional file 1: Fig. S4), indicating a good reproducibility of cell-specific transcriptomes.

By counting the raw transcript number of observed UMIs, we detected ~25,000 transcripts per egg cell and ~230,000 transcripts per central cell (Additional file 2: Table S1). Based on the abundance of all ERCC spike-in controls, we found that the capture efficiency of mRNA in scRNA-seq libraries varied from ~2 to ~7%. To eliminate the effect of capture efficiency variation and more importantly to compare expression



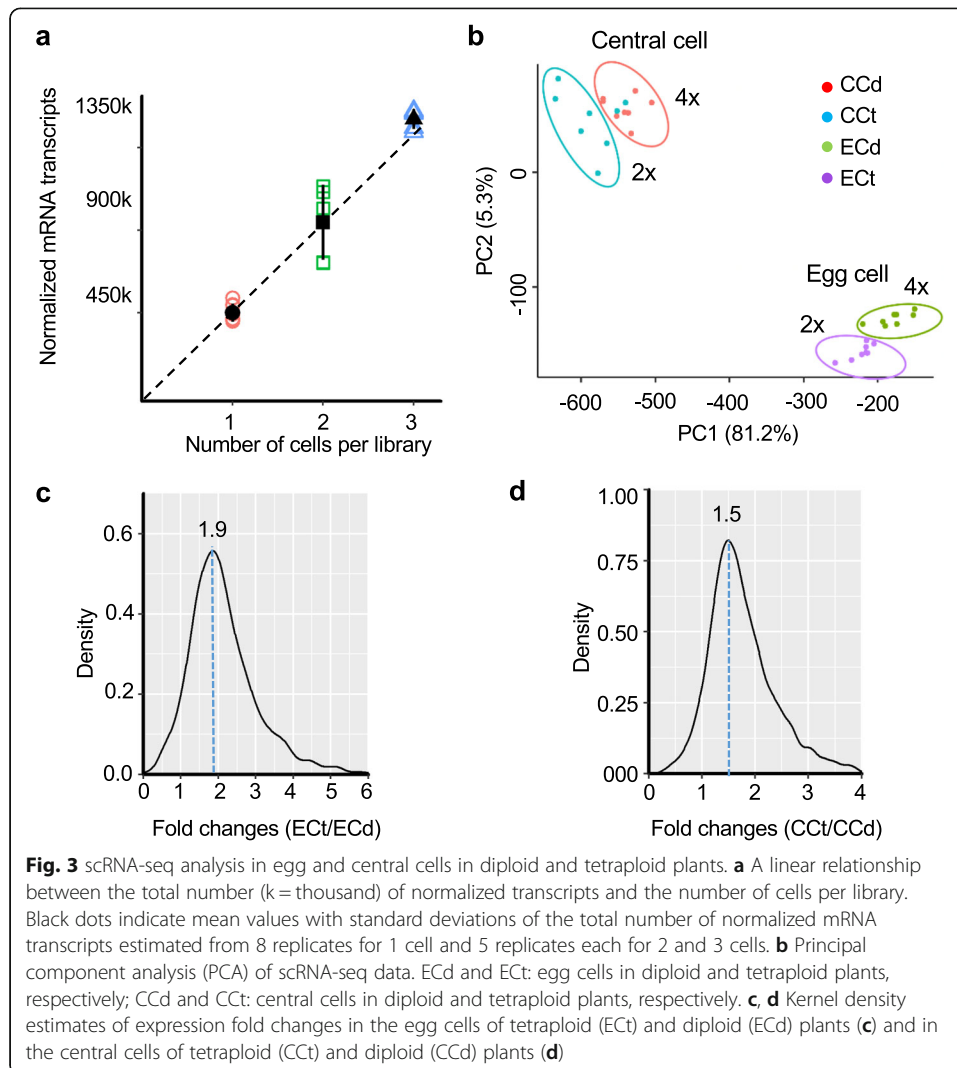
abundance among different cell types, we normalized average transcript abundance of each gene per egg or central cell using the capture efficiency (the “Methods” section). As expected, the dependence of observed molecule counts (Mol_{obs}) on expected molecule counts (Mol_{exp}) has fitted a Poisson generalized linear model [$\log_2(\text{Mol}_{\text{obs}}) = \beta_0 + \beta_1 \log_2(\text{Mol}_{\text{exp}})$] [42] (Additional file 1: Fig. S5). Thus, the capture efficiency and cell-to-cell technical noise were estimated using Poisson generalized linear regression of all ERCC spike-in controls for each cell (the “Methods” section). For example, the above raw numbers were normalized as average transcript abundance of $\sim 449,836$ per egg cell and $\sim 6,924,840$ per central cell. These normalized values were used for further analysis.

To confirm the reproducibility, we made additional RNA-seq libraries using an artificial mix of two or three egg cells in each library, each with five biological replicates. As expected, normalized numbers of total mRNA transcripts displayed a linear relationship with cell numbers per library (from $\sim 449,836$ per one egg cell to $\sim 1,507,362$ per three cells), in spite of some variation among replicates in the two-cell sample (Fig. 3a). The correlation between the expression levels and cell numbers was further analyzed by pairwise comparison of mRNA transcript number levels among the libraries containing different numbers of cells (Additional file 1: Fig. S6a, b). Linear regression analysis showed two and three cells per library had a 1.93-fold and 3.25-fold increase of transcript numbers, respectively, compared with one cell per library; highly expressed genes showed a better correlation between transcript abundance and cell numbers than poorly expressed genes. Despite relatively low power for single-cell analysis to detect genes that are expressed at low levels, this result validated a suitability of using scRNA-seq analysis for testing transcriptome changes between different cell types of diploid and tetraploid plants.

Transcriptome increase in the egg and central cells in tetraploid plants

Pearson’s correlation coefficients of scRNA-seq results among eight biological replicates were > 0.9 for egg cells and > 0.8 for central cells (Additional file 1: Fig. S6c), indicating low transcript number variation within the same type of female gametophytic cells. Principal component analysis (PCA) further separated the expression groups of all transcripts between cell types (egg and central cells) and between ploidy samples (diploid and tetraploid plants) (Fig. 3b). These results provided another valuable assessment of the quality and reproducibility of our scRNA-seq datasets (Additional file 2: Table S1).

For egg cells, normalized mRNA transcripts per cell were $\sim 449,836$ in the diploid plant (Additional file 1: Fig. S7a, b), which were doubled ($\sim 909,969$) in the tetraploid plant ($P = 0.0001$, Monte Carlo simulation, $N = 10,000$), indicating that genome duplication has a doubling effect on the transcriptome abundance in the egg cell. Consistent with the doubling effect, the peak of fold changes in the kernel density estimation was 1.9 between ECt and ECd (Fig. 3c). However, the central cell of the tetraploid plant showed ~ 1.6 -fold increase of the normalized transcript abundance (11,156,304) compared with that (6,924,840) in the diploid plant ($P = 0.0002$, Monte Carlo simulation, $N = 10,000$) (Additional file 1: Fig. S7c, d); the peak of fold changes was ~ 1.5 (Fig. 3d). At the genome-wide level, 57% and 58% of the expressed genes ($> 3,000$) showed larger than 1.5-fold increase after genome duplication in egg and central cells, respectively (Additional file 2: Table S1). The genes with higher expression levels tended to show more ploidy-dependent expression increase than the genes with lower expression levels

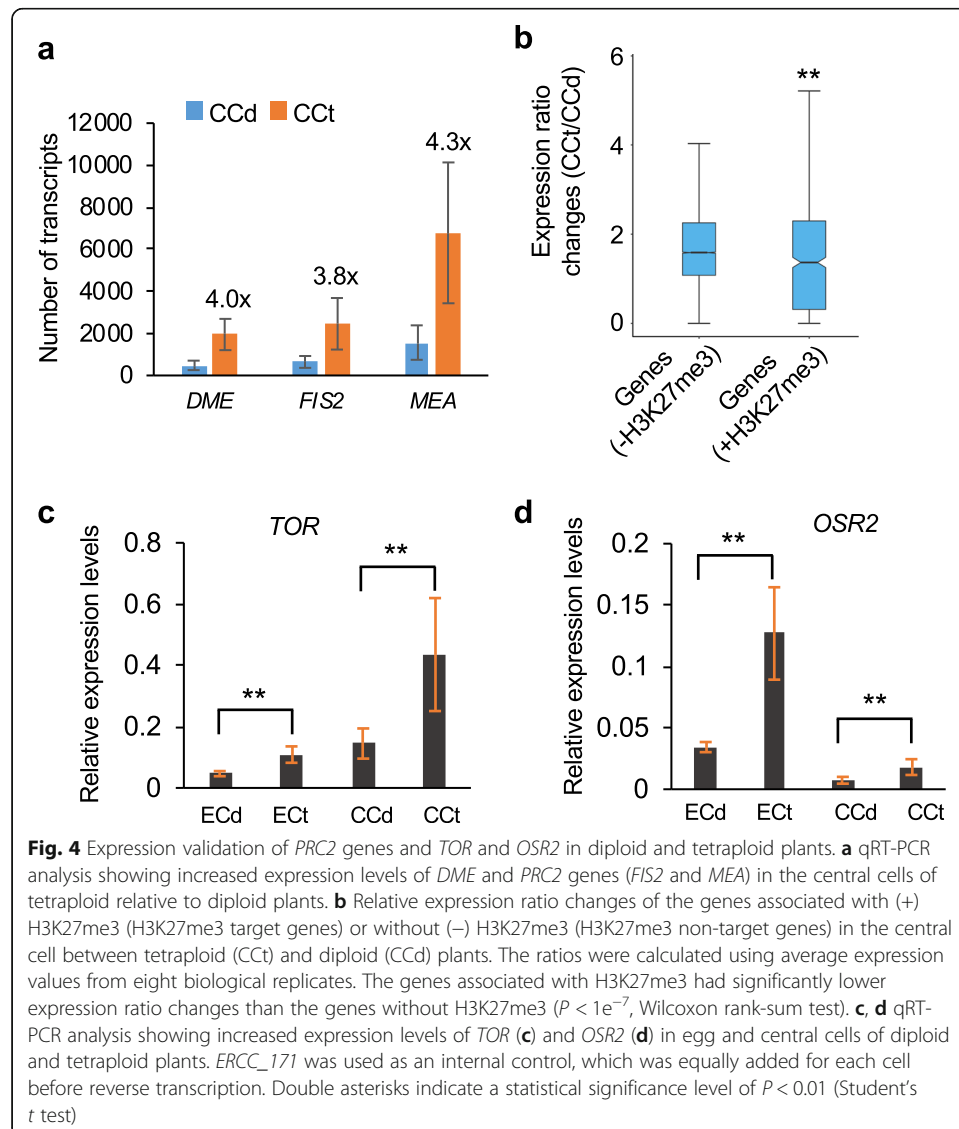


(Additional file 1: Fig. S7c, d). These results suggest that genome doubling in tetraploid plants can increase overall transcript and gene expression levels, but their fold increases depend on cell types.

Although the central cell has more transcripts than the egg cell, the fold increase of transcriptome changes in the tetraploids is smaller in the former (1.6-fold) than in the latter (2-fold). This could be related to genome-wide demethylation and de-repression of chromatin and/or transcriptional repressors in the central cell. Coincidentally, the central cell-specific DNA hypomethylation factor, *DEMETETER* (*DME*) [43], was expressed 4-fold higher in the tetraploid plant than in the diploid (Fig. 4a). *DME* activates Polycomb Repressive Complex 2 (*PRC2*) genes such as *fertilization-independent seed 2* (*FIS2*) [44] and *MEDEA* (*MEA*) [45] in the central cell [46]. As a result, expression levels of *FIS2* and *MEA* were increased from 3.8- to 4.3-fold in the tetraploid plant than in the diploid (Fig. 4a). Activation of *PRC2* genes can suppress the transcription of many other genes through the induction of histone H3K27 trimethylation (H3K27me3) [47, 48]. Indeed, H3K27me3 target genes identified in a previous study using the endosperm [49] showed significantly lower fold changes than the genes without H3K27me3

in the central cell of tetraploid plants ($P < 1e^{-7}$, Wilcoxon rank-sum test) (Fig. 4b). These results suggest that activation of *PRC2* genes may contribute to H3K27me3 increase and repression of overall expression levels in the central cell of tetraploid plants. Moreover, increased expression levels of *PRC2* genes in the central cell prior to fertilization could help bypass (endosperm) barriers of interspecific hybridization in allopolyploids as previously predicted [50, 51]. Notably, these genes (*DME*, *FIS2*, and *MEA*) are expressed specifically in the central cell and later in the endosperm but not in the egg cell [38] (Additional file 2: Table S1).

Among the upregulated genes by genome doubling, expression changes of 1,344 and 1,978 genes had larger than a 2-fold increase ($P < 0.05$, one-way ANOVA) in egg and central cells, respectively (Additional file 3: Table S2). Gene Ontology (GO) analysis showed enrichment of these genes in transcription- and translation-related processes, including translational initiation, mRNA processing, and protein folding (Additional file 1: Fig. S8). The larger size of the egg and central cells in the tetraploid than in the diploid led us to examine whether regulators of cell expansion were also upregulated after genome



duplication. *Arabidopsis* has several key factors involved in cell expansion and size, including ARGOS-LIKE (ARL) [52], TARGET OF RAPAMYCIN (TOR) [53], THESEUS1 (THE1) [54], and ORGAN SIZE RELATED 2 (OSR2) [55]. *ARL* and *THE1* showed very low or no expression in the egg and central cells of diploid plants (Additional file 2: Table S1). Interestingly, expression levels of *OSR2* and *TOR* in the egg and central cells were increased twofold or more in the tetraploid plants relative to the diploid ones (Additional file 2: Table S1); the result was further validated by qRT-PCR (Fig. 4c, d). These data suggest that the cell size regulators may also contribute to cell size increase in response to a ploidy increase.

Consistent with the increased cell size by ploidy, the central cell (diploid) is larger than the egg cell (haploid) from the same diploid plant, albeit of their different cell types (Fig. 2). The central cell had 6,924,840 normalized mRNA transcripts, 15-fold higher than the egg cell in the same diploid plant (Additional file 1: Fig. S7a, c). In addition, the synergid cell in the diploid plant had ~66,773 raw transcripts or ~2.9 million normalized mRNA transcripts. This suggests that the synergid cell contains more transcripts than the egg cell but less than the central cell, which is consistent with the published observation that the synergid cell is larger than the egg cell but smaller than the central cell [56]. These results suggest a positive correlation between RNA content and cell size [57]. The disproportionally larger than twofold increase in the transcriptome abundance among cell types probably reflects different activities in cell types; central cells are more metabolically active than the egg cells [58].

Cell-specific gene expression in egg, central, and synergid cells in diploid plants

Previous studies have generated transcriptome data using micro-dissected female gametophytes or sporocyteless mutant in *Arabidopsis* [59], wheat [60], rice [61], and maize [62]. In *Arabidopsis*, laser capture microdissection (LCM) combined with microarray or RNA-seq was commonly used to study gene expression changes in female gametophytic cells [63–65], which could result in datasets with mRNA cross-contamination among different cell types [66]. The quality of our scRNA-seq data was tested in the egg cell, central cell, and synergid cell (Fig. 5a). For comparative analysis, both scRNA-seq (this study) and published LCM-RNA-seq [64, 65] datasets were normalized by transcripts per million (TPM) (the “Methods” section). As expected, a subset of known gamete-specifically expressed genes (AT2G21740, AT1G74480, and AT2G21750 in EC; AT5G38330, AT3G10890, AT4G25530, AT5G04560, and AT3G04540 in CC; and AT1G47470, AT5G43510, AT4G18770, AT5G42955, AT4G07515, AT1G52970, AT2G21655, and AT5G12380 in SC) as reported [34, 41], also exhibited cell-specific expression patterns in our scRNA-seq datasets (Fig. 5a, upper panel). However, expression patterns of these genes were mixed among cell types in the LCM-RNA-seq datasets (Fig. 5a, upper right panel). The scRNA-seq datasets were largely contamination free, whereas LCM datasets included some seed coat genes (AT1G61720, AT5G48100, AT5G35550, AT4G09960, and AT5G17220) with low expression levels in female gametes.

In scRNA-seq datasets, we identified 5,456, 14,619, and 5,460 genes that were expressed (read count > 0 in any of the libraries) in the egg, central, and synergid cells, respectively (Fig. 5b). A total of 12,738 genes (39% of all expressed genes) showed significantly different expression patterns in at least two of the three (egg, central, and synergid) cell types ($P < 0.05$, one-way ANOVA), indicating enormous expression

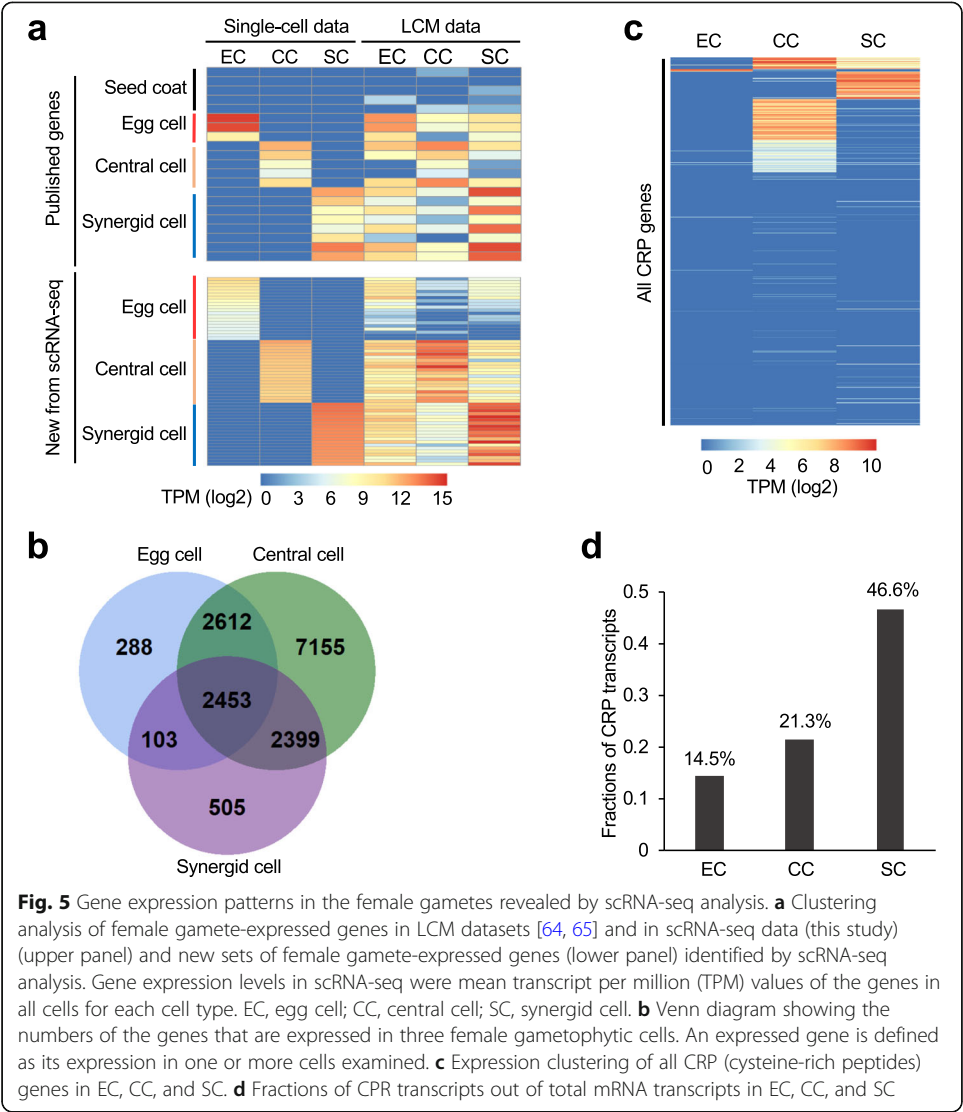


Fig. 5 Gene expression patterns in the female gametes revealed by scRNA-seq analysis. **a** Clustering analysis of female gamete-expressed genes in LCM datasets [64, 65] and in scRNA-seq data (this study) (upper panel) and new sets of female gamete-expressed genes (lower panel) identified by scRNA-seq analysis. Gene expression levels in scRNA-seq were mean transcript per million (TPM) values of the genes in all cells for each cell type. EC, egg cell; CC, central cell; SC, synergid cell. **b** Venn diagram showing the numbers of the genes that are expressed in three female gametophytic cells. An expressed gene is defined as its expression in one or more cells examined. **c** Expression clustering of all CRP (cysteine-rich peptides) genes in EC, CC, and SC. **d** Fractions of CPR transcripts out of total mRNA transcripts in EC, CC, and SC

variation among female gametes. Compared to the egg and synergid cells, the central cell had 2.7-fold more expressed genes, suggesting higher metabolic and biological activities in the central cell (Fig. 5b). Among those genes expressed in cell types, 288, 505, and 7,155 were exclusively expressed in the egg cell, synergid cell, and central cell, respectively (Fig. 5b), including top 20 genes with the highest expression values that are uniquely expressed in each cell type (Fig. 5a, lower left panel). The genes with expression patterns shared among cell types include 2,453 in all three cell types, 2,612 in EC and CC, 103 in EC and SC, and 2,399 in CC and SC (Fig. 5b). GO enrichment analysis of the female gamete-expressed genes indicated that some biological processes are shared among three female cell types, while others are specific to each cell type: embryo development in the egg cell, photosynthesis and response to cytokinins in the central cell, and pollen germination and small GTPase-mediated signal transduction in the synergid cell (Additional file 1: Fig. S9). These results are consistent with respective biological roles in three cell types: embryo development in the egg cell, photosynthesis and

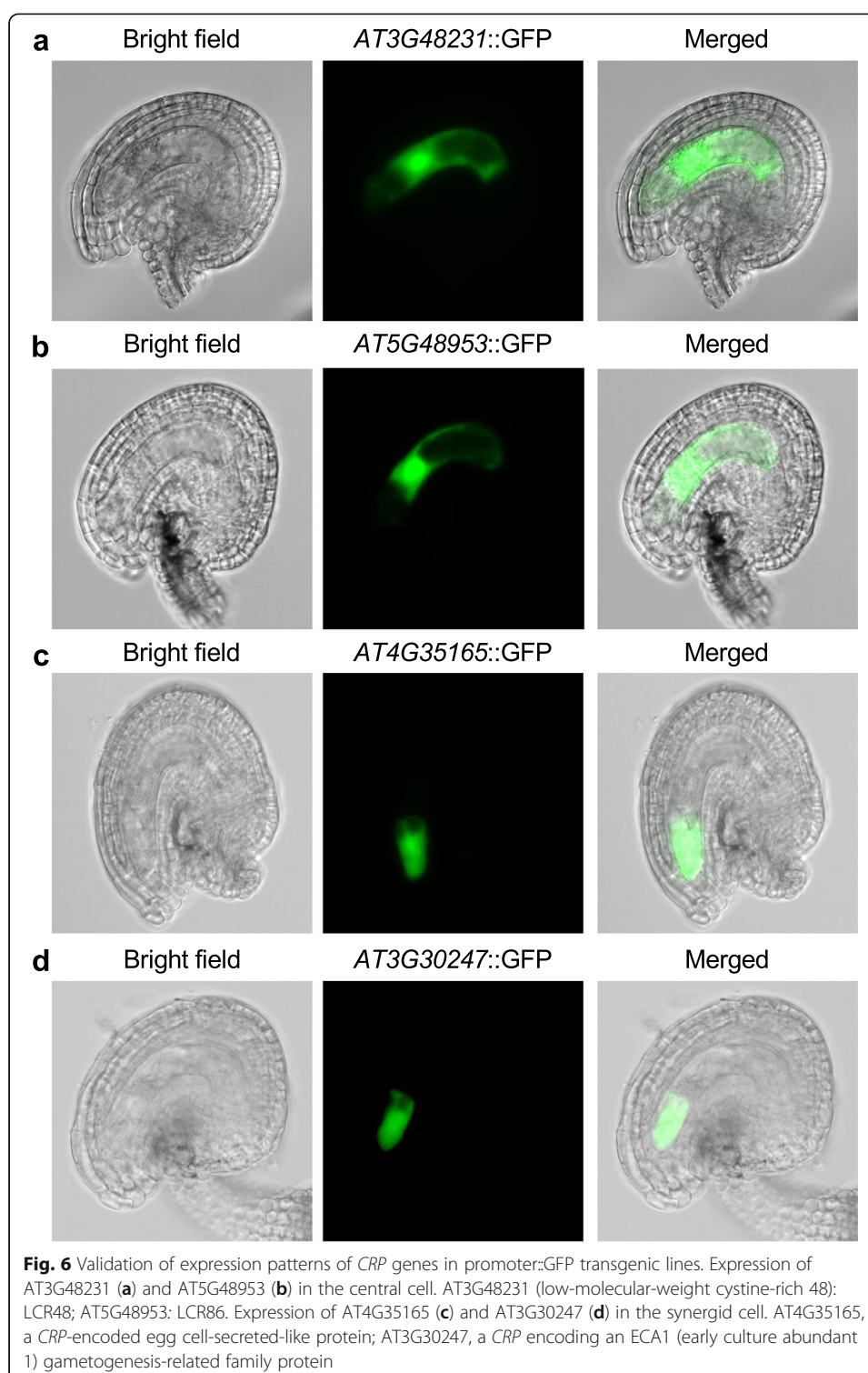
energy metabolism in the central cell (progenitor of the endosperm), and assisting pollen tube growth and fertilization in the synergid cell.

Remarkably, a group of gene family members encoding cysteine-rich peptides (CRPs) [67], accounted for a large amount of total normalized mRNA transcripts in the egg (15%), central (21%), and synergid (46%) cells (Fig. 5c, d). CRPs regulate plant growth and development through modulation of cell-cell communications, including the guidance of pollen tube and gamete recognition during fertilization [58, 68]. Among these CRP genes (Additional file 4: Table S3), we randomly selected four newly defined genes (two expressed in the synergid cell and two expressed in the central cell) for functional validation by expressing promoter::GFP in the transgenic lines (Fig. 6). Consistent with the scRNA-seq data, two genes (AT5G48953 and AT3G48231) were expressed in the central cell (Fig. 6a, b), according to the image evaluation methods as previously described [34], while the other two (AT4G35165 and AT3G30247) were expressed in the synergid cell (Fig. 6c, d). The data suggest roles of these cell-specifically expressed genes in female-male gametic interactions between synergid cell and pollen tube (sperm) for pollen guidance and between central cell and sperm in double fertilization for endosperm development.

Discussion

Single-cell analysis provides new opportunities to study complex cellular systems at the individual cell resolution. It is used not only to dissect cell heterogeneity in given tissues, but also to quantify absolute expression level in each cell. In this study, we have employed scRNA-seq to examine the relationships between ploidy, cell size, and transcriptome abundance. Cell types and sizes vary from one tissue to another in plants. For example, the commonly used leaf cells are often mixed with diploid and endoreduplicated cells [16], which could confound the effects of polyploidy (WGD) and endoreduplication [22]. Although cell types are well-defined in root cells [69], they show large gene expression variation from one cell type to another from > 26-fold in the epidermal cell to > 3-fold in the quiescent center cell [25]. In our study, cell size and expression values of female gametophytic cells show less variation from one plant to another in similar stages at the same ploidy level (Fig. 2 and Additional file 1: Fig. S7). This is probably because the synchrony of gametes is critical to fertilization and embryo development. For example, cell cycle regulation is coordinated between male and female gametes for double fertilization [70]; both egg and central cells are arrested at the G1/S transition in tobacco [71].

Our data indicate that in the tetraploid plants relative to the diploids, overall transcript abundance per cell is doubled in the egg cell and increased in the central cell, which is consistent with the cell size increase in these cells. This conclusion is different from previously published results in which fewer genes are upregulated in response to ploidy increase in yeast ploidy series [17] and in *Arabidopsis* autotetraploids [19]. When cells are sorted by the nuclei of diploid and endoreduplicated cells in tomato pericarps, the transcriptome abundance is correlated with the ploidy level of the cells [22]. Cell size is positively correlated with transcriptome abundance, as observed between transcriptome abundance and cell size in yeast, plant, and mammals [21, 72]. In yeast, genes controlling cell cycles, *Cln1* and *Pcl1*, are repressed as the ploidy levels increase, leading to larger cell size [17]. In *Arabidopsis*, because these female gametic cells do



not divide, the cell size increase is likely related to cell expansion [73, 74]. This cell expansion regulation involves several key factors, including *OSR2* [55] and its family member *ARL* [52], *TOR* kinase [53], and receptor-like kinase *THESEUS1* (*THE1*) [54]. *TOR* kinase is evolutionarily conserved and functions to mediate ribosomal biogenesis and translation that promote cell proliferation and growth including embryogenesis

and life span determination [75]. Ectopic expression of *TOR* [53] or *OSR2* [55] results in larger cells than the wild type, while downregulation of *ARL* inhibits cell growth [52]. *THE1* is required for cell elongation through brassinosteroid-mediated signaling, and downregulation or mutation of *THE1* results in dwarf plants [54]. Thus, in addition to the polyploid effect, cell size regulators such as *TOR* [53] or *OSR2* [55] may also affect cell size increase in the egg and central cells. Increased expression of these genes in the egg and central cell of tetraploid plants relative to the diploids suggest their potential roles in cell size variation, although the conclusion for a causal effect should await further experimentation.

The variation of transcriptome abundance increase between the cell types is likely associated with the biological activity of each cell type. In egg cells that maintain a quiescent state, genome doubling increases the cell size and doubles the transcriptome. In the central cell, epigenetic reprogramming by the activation of *DME*, which is not expressed in the egg cell, could alter the overall impact of transcriptome changes. Interestingly, *DME* [45] is upregulated in the central cell of tetraploid plants relative to the diploids (Fig. 4a); this could activate general transcriptional repressors such as *PRC2* gene families including *FIS2* [44] and *MEA* [45] that can promote H3K27 trimethylation of many *PRC2* target genes in the central cell [47, 48]. Consistently, the expression fold change of H3K27me3 target genes was significantly lower than that of H3K27me3 non-target genes in the central cell between tetraploid (CCt) and diploid (CCd) plants (Fig. 4b). As a result, the overall increase of transcript abundance is not exactly to the twofold ploidy level as observed in the egg cell. Although this prediction is based on the correlated data of gene expression, it can be tested by examining the transcriptional activity of egg and central cells in the *dme* mutant. Unfortunately, in spite of concerted and collaborative efforts, we could not isolate intact central cells in the *dme* line, and the central cells were easily broken with dissipated nuclei. It is likely that the disruption of *DME* or *PRC2* complex genes may affect the cellular morphology and make the central cells sensitive to physical manipulations.

Compared to previous studies, our scRNA-seq datasets are largely free of cross-contamination among three female gametophytic cells, which provide a unique resource for studying plant reproductive biology. Many genes with cell-specific expression patterns in each female gamete identified in this study can provide useful clues for better understanding the molecular events of cell-cell recognition during fertilization. CRPs are involved in diverse aspects of cell-cell communication during vegetative growth and plant reproduction [58, 68]. As the amino acid composition of CRPs is highly divergent, different CRPs may play specific and unique functions in the egg cell, central cell, or synergid cell during fertilization. For example, CRPs secreted from synergid cells can provide guidance for pollen tube and sperm cell release, while those from the central cell and egg cells can attract sperm to produce embryo and endosperm of a seed [58, 71], a process known as double fertilization [70, 76]. These newly identified female gamete-expressed CRPs are key signaling factors in regulating reproductive development in plants. Compared to the egg cell, the function of CRPs in the central cell and synergid cell is poorly understood. Our scRNA-seq resource should provide valuable guidelines for future research directions to elucidate the roles of egg, central, and synergid cells during double fertilization [58, 68]. Understanding the molecular mechanisms for the polyploid effects on cell and organ size and for signaling processes

Methods

A. thaliana (Col-0) diploids and tetraploids were obtained by colchicine treatment and confirmed by chromosome spreads and flow cytometry [16]. The same batch of seed stocks was used for this study. Diploid and tetraploid plants were independently transformed with each of the constructs. Transgenic plants (*A. thaliana* Col-0) showing cell-specific promoter:nGFP of pDD45:nGFP (in egg cell nucleus) [34], pSUP16:nGFP (in central cell nucleus) [35], or pDD31:nGFP (in synergid cell nucleus) [35] were generated in *A. thaliana* (ecotype Col-0) diploid and isogenic tetraploids [16]. All transgenic plants were grown under the long-day condition (light/dark cycle of 16/8 h) at 22 °C. Flowers at stage 12 were tagged and emasculated, and the emasculated plants were grown for another day. The ovules were dissected from pistils and soaked in a protoplast enzyme solution (2% cellulose, 0.3% macerozyme R-10, 0.05% pectolyase, and 0.45 M mannitol) for 15 min as previously described [77]. Individual egg, central, or synergid cells labeled independently by nucleus-localized GFP were carefully isolated from different plants in various times and captured into a microtube using a microcapillary pipette that is installed in a micromanipulator (MN-151, NARISHIGE) and a micro-injector (IM-11-2, NARISHIGE) under an inverted dissecting microscope (Eclipse Ts2R, Nikon).

scRNA-seq libraries were constructed using a modified protocol as previously described [26, 36]. Each single cell was placed into 4 μ l lysis buffer (0.1% Triton X-100, 1 U/ μ l RNase-OUT Ribonuclease Inhibitor (Thermo Fisher Scientific, Waltham, MA), 4 μ M oligo-dT primer (5'-GTTTCAGAGTTCTACAGTCCGACGATCGT'TTTTTTTTTTTTTTTTTTTTTTTTTTTTTT-3'), 2.5 mM dNTP, and 1 μ l 1:3,000,000 ERCC Spike-In Mix 1 (Thermo Fisher Scientific, Waltham, MA) with ~ 20,775 RNA molecules). Samples were incubated at 72 °C for 3 min and immediately placed on ice. After lysis, each sample was added by 5.6 μ l of RT mix, consisting of 2 μ l Superscript II first-strand buffer (5 \times , Thermo Fisher Scientific), 0.5 μ l DTT (100 mM, Thermo Fisher Scientific), 2 μ l Betaine (5 M, Sigma-Aldrich, St. Louis, MO), 0.1 μ l MgCl₂ (1 M, Sigma-Aldrich), 0.25 μ l TSO primer (100 μ M, 5'-GUUCAGAGUUCUACAGUCCGACGAUC HNNNNNGGG-3'), 0.25 μ l RNaseOUT Ribonuclease Inhibitor, and 0.5 μ l SuperScript II reverse transcriptase (200 U/ μ l, Thermo Fisher Scientific). Reverse transcription reaction was performed by incubating samples at 42 °C for 90 min, followed by 10 cycles of reverse transcription (50 °C for 2 min and 42 °C for 2 min) and one cycle of extension (72 °C for 15 min). The cDNA was amplified by adding 15 μ l PCR mix (12.5 μ l KAPA HiFi HotStart ReadyMix (2 \times , Kapa Biosystems, Wilmington, MA), 0.25 μ l IS PCR primer (10 μ M, 5'-GTTTCAGAGTTCTACAGTCCGACGATC-3'), and 2.25 μ l water) and incubating at a thermal cycler as follows: 98 °C for 3 min; 18–20 cycles of 98 °C for 20 s, 68 °C for 15 s, and 72 °C for 6 min; and 72 °C for 5 min. PCR product was purified using AMPure XP beads (1:1 ratio; Beckman Coulter) and then used for qRT-PCR or library construction.

For qRT-PCR, 100 pg amplified DNA was used as the template, and the reaction was run on the LightCycler 96 System (Roche, Indianapolis, ID). The relative expression

level was quantified using internal control *ERCC_171* with six biological replicates, and three technical replicates were used for each biological replicate. Primers for qRT-PCR were listed in Additional file 5: Table S4.

For library construction, 200 pg amplified DNA was tagged using Nextera XT DNA sample preparation kit according to the manufacturer's instructions (Illumina, San Diego, CA). The fragmented DNA was purified using AMPure XP beads (1:1 ratio; Beckman Coulter) and incubated with PvuI-HF enzyme (NEB, Ipswich, MA) for 30 min at 37 °C. After digestion by PvuI-HF, DNA was purified using AMPure XP beads (1:1 ratio; Beckman Coulter) and suspended in 18 µl water. Final amplification was performed by adding 22 µl PCR mix (20 µl KAPA HiFi HotStart ReadyMix (2×, Kapa Biosystems, Wilmington, MA), 1 µl FP PCR primer (10 µM, 5'-AATGATACGGCGACCACCGAGATCTACACGTTTCAGAGTTCTACAGTCCGA-3'), and 1 µl RP PCR primer (10 µM, 5'-CAAGCAGAAGACGGCATACGAGATXXXXXXGTCTCGTGGGCTCGG-3' (XXXXXX indicates index sequence) and incubating samples at a thermal cycler as follows: 98 °C for 3 min; 12–15 cycles of 98 °C for 20 s, 65 °C for 30 s, and 72 °C for 1 min; and 72 °C for 5 min. Final libraries were purified using AMPure XP beads (0.8:1 ratio; Beckman Coulter) and sequenced using NextSeq 500 platform (Single-end 75-bp reads).

Read mapping of scRNA-seq datasets

We first removed low-quality reads that showed a quality score of 29 or lower in the first six bases using the NGS QC Toolkit (version 2.3) and then discarded reads without the UMI pattern HNNNNNGGG at 5' ends. The remaining reads were mapped to the pseudogenome consisting of *Arabidopsis* genome (TAIR 10) and ERCC sequences using STAR with parameters (--outFilterMismatchNoverLmax 0.02), and non-uniquely mapped reads were discarded [78]. To account for the incomplete information of transcription start sites, the 5' ends of all gene models were extended by 100 bases but not beyond the 3' ends of upstream genes. Each unique UMI barcode represents one RNA transcript. Errors in the UMI sequence generated during PCR and sequencing could create additional artificial UMIs. To account for sequencing errors and remove PCR duplicates, UMI-tools were applied to improve the quantification accuracy [79]. Molecule count of each gene was calculated as the total number of distinct UMIs mapped to the corresponding gene model.

Normalization of average transcript abundance in each scRNA-seq library

The dependence of observed molecule counts (Mol_{obs}) on expected molecule counts (Mol_{exp}) fitted a Poisson generalized linear model for ERCC spike-in controls (Additional file 1: Fig. S5). To account for capture efficiency variation and technical noise between different cells, we first fitted a Poisson generalized linear model (GLM) using the ERCC spike-in controls:

$$\log_2(Mol_{obs}) = \beta_0 + \beta_1 \log_2(Mol_{exp})$$

where Mol_{exp} is the expected ERCC molecule counts, and Mol_{obs} is the observed raw ERCC molecule counts [42]. The slope (β_1) and intercept (β_0) of ERCC spike-in controls in each library were calculated using the *glm* function in R software as “*glm(y~log2(x), family = 'Poisson')*” in which x and y were the Mol_{exp} and Mol_{obs} , respectively, of the full set of ERCC spike-in transcripts. Endogenous genes were expected to show a similar relationship between observed raw molecule counts and actual molecule counts

as ERCC spike-in controls. Then, we transformed observed molecule counts of endogenous genes ($\text{Mol}_{\text{obs_gene}}$) with slope (β_1) and intercept (β_0) of the Poisson GLM regression line to generate the normalized expression levels of genes ($\text{Mol}_{\text{norm_gene}}$) for each cell:

$$\log_2(\text{Mol}_{\text{norm_gene}}) = (\log_2(\text{Mol}_{\text{obs_gene}}) - \beta_0) / \beta_1$$

Identification of upregulated genes by genome doubling

Data in each cell was treated as a biological replicate. Normalized expression abundance of each gene with eight replicates was subjected to a one-way ANOVA test using *f_oneway* function in the SciPy software (<https://www.scipy.org>) to identify upregulated genes between ECt and ECd or between CCt and CCd with a statistical significance level of $P < 0.05$.

Principal component analysis and Gene Ontology analysis

Normalized expression abundance of genes in egg cells and central cells from diploid and tetraploid plants was subject to logarithm transformation using *log2* function in R software. PCA was performed using the *prcomp* function in R software with the parameters “scale. = FALSE, center = FALSE, tol = 0”.

GO analysis was performed using the Database for Annotation, Visualization and Integrated Discovery (DAVID) v6.8 (<https://david.ncifcrf.gov/home.jsp>) [80]. The enriched biological process terms were extracted from the GOTERM_BP_DIRECT category with Benjamini-corrected P value < 0.05 .

Analysis of published LCM datasets and genes expressed in female gametes

Raw reads from LCM datasets were mapped to *Arabidopsis* genome (TAIR 10) using the HISAT2 (v2.1.0) software with default parameters [81]. Uniquely mapped reads were extracted and used to calculate transcripts levels (transcripts per million (TPM)) of each gene through the StringTie (v1.3.3) software [82].

To compare our scRNA-seq datasets with the LCM datasets, we calculated the TPM value of each gene in scRNA-seq datasets. For each gene, the total number of distinct UMIs was calculated as the transcript number. The sum of transcript numbers of all genes was divided by 1,000,000 to create a scaling factor. Transcript number of each gene is divided by the scaling factor to generate the TPM value of each gene.

Expression validation of CRP genes

Genomic DNA from *Arabidopsis* (Col-0) was used to amplify promoter regions of *AT3G4823*, *AT5G48953*, *AT4G35165*, and *AT3G30247*. Primer pairs are listed in Additional file 5: Table S4. The amplified fragments were cloned into pBI-n1GFP vector. All constructs were individually cloned into *Agrobacterium* strain GV3101 and then transformed into diploid *Arabidopsis* (Col-0) using the standard floral dip method [83]. GFP activity within the mature female gametophyte was captured 1 day after emasculation using Zeiss Axiovert 200M fluorescence microscope.

Supplementary information

Supplementary information accompanies this paper at <https://doi.org/10.1186/s13059-020-02094-0>.

Additional file 1: Supplementary Figures S1–S9.

Additional file 2: Table S1. Expression values (raw reads and normalized abundance) of each gene in each cell.

Additional file 3: Table S2. List of differentially expressed genes after genome duplication at a statistical significance level ($P < 0.05$, one-way ANOVA) using normalized expression values.

Additional file 4: Table S3. Normalized expression values of each CRP gene in each cell.

Additional file 5: Table S4. Primer pairs used for qRT-PCR and GFP validation.

Additional file 6: Review history.

Acknowledgements

We thank Chad Williams for the assistance with library construction, Changqing Zhang for the GFP transgenic lines, and the Genomic Sequencing and Analysis Facility at The University of Texas at Austin for the high-throughput sequencing.

Peer review information

Kevin Pang was the primary editor of this article and managed its editorial process and peer review in collaboration with the rest of the editorial team.

Review history

The review history is available as Additional file 6.

Authors' contributions

Q.S., A. A., and Z.J.C. conceived the research, analyzed the data, and wrote the paper. Q.S. and A.A. performed the experiments. N.J.J. and Y.I. provided materials, reagents, and technical assistance. The authors read and approved the final manuscript.

Funding

The work is supported by grants to Z.J.C. from the National Institutes of Health (GM109076) and the National Science Foundation (IOS1739092) and grant to Q.S. from the Natural Science Foundation of Jiangsu Province (BK20190510).

Availability of data and materials

scRNA-seq data were deposited in the NCBI Nucleotide and Sequence Read Archive (SRA) under Accession SRP160651 [84]. All codes used for statistical analysis are available upon request or at GitHub [85].

Ethics approval and consent to participate

Not applicable.

Competing interests

The authors declare no competing interests.

Author details

¹Department of Molecular Biosciences, The University of Texas at Austin, 1 University Station A5000, Austin, TX 78712, USA. ²State Key Laboratory of Crop Genetics and Germplasm Enhancement, Nanjing Agricultural University, Nanjing 210095, China. ³Department of Biomedical Engineering, The University of Texas at Austin, 1 University Station C0800, Austin, TX 78712, USA. ⁴Institute of Plant Science and Resources, Okayama University, Chuo 2-20-1, Kurashiki, Okayama 710-0046, Japan.

Received: 27 February 2020 Accepted: 6 July 2020

Published online: 22 July 2020

References

- Wendel JF, Jackson SA, Meyers BC, Wing RA. Evolution of plant genome architecture. *Genome Biol.* 2016;17:37.
- Comai L. The advantages and disadvantages of being polyploid. *Nat Rev Genet.* 2005;6:836–46.
- Soltis DE, Visger CJ, Soltis PS. The polyploidy revolution then...and now: Stebbins revisited. *Am J Bot.* 2014;101:1057–78.
- Adams KL, Wendel JF. Polyploidy and genome evolution in plants. *Curr Opin Plant Biol.* 2005;8:135–41.
- Jackson S, Chen ZJ. Genomic and expression plasticity of polyploidy. *Curr Opin Plant Biol.* 2010;13:153–9.
- Leitch AR, Leitch IJ. Genomic plasticity and the diversity of polyploid plants. *Science.* 2008;320:481–3.
- Storchova Z, Pellman D. From polyploidy to aneuploidy, genome instability and cancer. *Nat Rev Mol Cell Biol.* 2004;5:45–54.
- Bevan MW, Uauy C, Wulff BB, Zhou J, Krasileva K, Clark MD. Genomic innovation for crop improvement. *Nature.* 2017; 543:346–54.
- Chen ZJ, Sreedasyam A, Ando A, Song Q, De Santiago LM, Hulse-Kemp AM, Ding M, Ye W, Kirkbride RC, Jenkins J, et al. Genomic diversifications of five *Gossypium* allopolyploid species and their impact on cotton improvement. *Nat Genet.* 2020;52:525–33.
- Selmecki AM, Maruvka YE, Richmond PA, Guillet M, Shores N, Sorenson AL, De S, Kishony R, Michor F, Dowell R, Pellman D. Polyploidy can drive rapid adaptation in yeast. *Nature.* 2015;519:349–52.
- Chao DY, Dillkes B, Luo H, Douglas A, Yakubova E, Lahner B, Salt DE. Polyploids exhibit higher potassium uptake and salinity tolerance in *Arabidopsis*. *Science.* 2013;341:658–9.
- Ni Z, Kim ED, Ha M, Lackey E, Liu J, Zhang Y, Sun Q, Chen ZJ. Altered circadian rhythms regulate growth vigour in hybrids and allopolyploids. *Nature.* 2009;457:327–31.

13. Miller M, Song Q, Shi X, Juenger TE, Chen ZJ. Natural variation in timing of stress-responsive gene expression predicts heterosis in intraspecific hybrids of *Arabidopsis*. *Nat Commun*. 2015;6:7453.
14. Song Q, Ando A, Xu D, Fang L, Zhang T, Huq E, Qiao H, Deng XW, Chen ZJ. Diurnal down-regulation of ethylene biosynthesis mediates biomass heterosis. *Proc Natl Acad Sci U S A*. 2018;115:5606–11.
15. Chen ZJ. Genomic and epigenetic insights into the molecular bases of heterosis. *Nat Rev Genet*. 2013;14:471–82.
16. Miller M, Zhang C, Chen ZJ. Ploidy and hybridity effects on growth vigor and gene expression in *Arabidopsis thaliana* hybrids and their parents. G3 (Bethesda). 2012;2:505–13.
17. Galitski T, Saldanha AJ, Styles CA, Lander ES, Fink GR. Ploidy regulation of gene expression. *Science*. 1999;285:251–4.
18. Coate JE, Doyle JJ. Quantifying whole transcriptome size, a prerequisite for understanding transcriptome evolution across species: an example from a plant allopolyploid. *Genome Biol Evol*. 2010;2:534–46.
19. Yu Z, Haberer G, Matthes M, Rattei T, Mayer KF, Gierl A, Torres-Ruiz RA. Impact of natural genetic variation on the transcriptome of autotetraploid *Arabidopsis thaliana*. *Proc Natl Acad Sci U S A*. 2010;107:17809–14.
20. Wang J, Tian L, Lee HS, Wei NE, Jiang H, Watson B, Madlung A, Osborn TC, Doerge RW, Comai L, Chen ZJ. Genomewide nonadditive gene regulation in *Arabidopsis* allotetraploids. *Genetics*. 2006;172:507–17.
21. Coate JE, Doyle JJ. Variation in transcriptome size: are we getting the message? *Chromosoma*. 2015;124:27–43.
22. Pirrello J, Deluche C, Frangne N, Gevaudant F, Maza E, Djari A, Bourge M, Renaudin JP, Brown S, Bowler C, et al. Transcriptome profiling of sorted endoreduplicated nuclei from tomato fruits: how the global shift in expression ascribed to DNA ploidy influences RNA-Seq data normalization and interpretation. *Plant J*. 2018;93:387–98.
23. Loven J, Orlando DA, Sigova AA, Lin CY, Rahl PB, Burge CB, Levens DL, Lee TI, Young RA. Revisiting global gene expression analysis. *Cell*. 2012;151:476–82.
24. Birchler JA. Facts and artifacts in studies of gene expression in aneuploids and sex chromosomes. *Chromosoma*. 2014;123:459–69.
25. Brennecke P, Anders S, Kim JK, Kolodziejczyk AA, Zhang X, Proserpio V, Baying B, Benes V, Teichmann SA, Marioni JC, Heisler MG. Accounting for technical noise in single-cell RNA-seq experiments. *Nat Methods*. 2013;10:1093–5.
26. Islam S, Zeisel A, Joost S, La Manno G, Zajac P, Kasper M, Lonnberg P, Linnarsson S. Quantitative single-cell RNA-seq with unique molecular identifiers. *Nat Methods*. 2014;11:163–6.
27. Karaïskos N, Wahle P, Alles J, Boltengagen A, Ayoub S, Kipar C, Kocks C, Rajewsky N, Zinzen RP. The *Drosophila* embryo at single-cell transcriptome resolution. *Science*. 2017;358:194–9.
28. Navin N, Kendall J, Troge J, Andrews P, Rodgers L, McIndoo J, Cook K, Stepansky A, Levy D, Esposito D, et al. Tumour evolution inferred by single-cell sequencing. *Nature*. 2011;472:90–4.
29. Efroni I, Birnbaum KD. The potential of single-cell profiling in plants. *Genome Biol*. 2016;17:65.
30. Libault M, Pingault L, Zogli P, Schiefelbein J. Plant systems biology at the single-cell level. *Trends Plant Sci*. 2017;22:949–60.
31. Doyle JJ, Coate JE. Polyploidy, the nucleotype, and novelty: the impact of genome doubling on the biology of the cell. *Int J Plant Sci*. 2019;180:1–52.
32. Ng DW, Zhang C, Miller M, Shen Z, Briggs SP, Chen ZJ. Proteomic divergence in *Arabidopsis* autopolyploids and allopolyploids and their progenitors. *Heredity*. 2012;108:419–30.
33. Hu Z, Chen K, Xia Z, Chavez M, Pal S, Seol JH, Chen CC, Li W, Tyler JK. Nucleosome loss leads to global transcriptional up-regulation and genomic instability during yeast aging. *Genes Dev*. 2014;28:396–408.
34. Steffen JG, Kang IH, Macfarlane J, Drews GN. Identification of genes expressed in the *Arabidopsis* female gametophyte. *Plant J*. 2007;51:281–92.
35. Wang D, Zhang C, Hearn DJ, Kang IH, Punwani JA, Skaggs MI, Drews GN, Schumaker KS, Yadegari R. Identification of transcription-factor genes expressed in the *Arabidopsis* female gametophyte. *BMC Plant Biol*. 2010;10:110.
36. Picelli S, Faridani OR, Bjorklund AK, Winberg G, Sagasser S, Sandberg R. Full-length RNA-seq from single cells using Smart-seq2. *Nat Protoc*. 2014;9:171–81.
37. Grun D, Kester L, van Oudenaarden A. Validation of noise models for single-cell transcriptomics. *Nat Methods*. 2014;11:637–40.
38. Choi Y, Gehring M, Johnson L, Hannon M, Harada JJ, Goldberg RB, Jacobsen SE, Fischer RL. DEMETER, a DNA glycosylase domain protein, is required for endosperm gene imprinting and seed viability in *Arabidopsis*. *Cell*. 2002;110:33–42.
39. Kinoshita T, Miura A, Choi Y, Kinoshita Y, Cao X, Jacobsen SE, Fischer RL, Kakutani T. One-way control of FWA imprinting in *Arabidopsis* endosperm by DNA methylation. *Science*. 2004;303:521–3.
40. Koszegi D, Johnston AJ, Rutten T, Czihal A, Altschmied L, Kumlehn J, Wust SE, Kirioukhova O, Gheyselinck J, Grossniklaus U, Baumelein H. Members of the RKD transcription factor family induce an egg cell-like gene expression program. *Plant J*. 2011;67:280–91.
41. Sprunck S, Rademacher S, Vogler F, Gheyselinck J, Grossniklaus U, Dresselhaus T. Egg cell-secreted EC1 triggers sperm cell activation during double fertilization. *Science*. 2012;338:1093–7.
42. Tung PY, Blischak JD, Hsiao CJ, Knowles DA, Burnett JE, Pritchard JK, Gilad Y. Batch effects and the effective design of single-cell gene expression studies. *Sci Rep*. 2017;7:39921.
43. Gehring M, Huh JH, Hsieh TF, Penterman J, Choi Y, Harada JJ, Goldberg RB, Fischer RL. DEMETER DNA glycosylase establishes MEDEA polycomb gene self-imprinting by allele-specific demethylation. *Cell*. 2006;124:495–506.
44. Chaudhury AM, Ming L, Miller C, Craig S, Dennis ES, Peacock WJ. Fertilization-independent seed development in *Arabidopsis thaliana*. *Proc Natl Acad Sci U S A*. 1997;94:4223–8.
45. Grossniklaus U, Vielle-Calzada JP, Hoepfner MA, Gagliano WB. Maternal control of embryogenesis by MEDEA, a polycomb group gene in *Arabidopsis*. *Science*. 1998;280:446–50.
46. Xiao W, Gehring M, Choi Y, Margossian L, Pu H, Harada JJ, Goldberg RB, Pennell RI, Fischer RL. Imprinting of the MEA Polycomb gene is controlled by antagonism between MET1 methyltransferase and DME glycosylase. *Dev Cell*. 2003;5:891–901.
47. Raissig MT, Berner M, Baroux C, Grossniklaus U. Genomic imprinting in the *Arabidopsis* embryo is partly regulated by PRC2. *PLoS Genet*. 2013;9:e1003862.
48. Kohler C, Wolff P, Spillane C. Epigenetic mechanisms underlying genomic imprinting in plants. *Annu Rev Plant Biol*. 2012;63:331–52.
49. Weinhofer I, Hehenberger E, Roszak P, Hennig L, Kohler C. H3K27me3 profiling of the endosperm implies exclusion of polycomb group protein targeting by DNA methylation. *PLoS Genet*. 2010;6:e1001152.
50. Bushell C, Spielman M, Scott RJ. The basis of natural and artificial postzygotic hybridization barriers in *Arabidopsis* species. *Plant Cell*. 2003;15:1430–42.
51. Josefsson C, Dilkes B, Comai L. Parent-dependent loss of gene silencing during interspecies hybridization. *Curr Biol*. 2006;16:1322–8.

52. Hu Y, Poh HM, Chua NH. The Arabidopsis ARGOS-LIKE gene regulates cell expansion during organ growth. *Plant J.* 2006;47:1–9.
53. Deprost D, Yao L, Sormani R, Moreau M, Leterreux G, Nicolai M, Bedu M, Robaglia C, Meyer C. The Arabidopsis TOR kinase links plant growth, yield, stress resistance and mRNA translation. *EMBO Rep.* 2007;8:864–70.
54. Guo H, Li L, Ye H, Yu X, Algreen A, Yin Y. Three related receptor-like kinases are required for optimal cell elongation in *Arabidopsis thaliana*. *Proc Natl Acad Sci U S A.* 2009;106:7648–53.
55. Qin Z, Zhang X, Zhang X, Feng G, Hu Y. The Arabidopsis ORGAN SIZE RELATED 2 is involved in regulation of cell expansion during organ growth. *BMC Plant Biol.* 2014;14:349.
56. Enghart M, Soljic L, Sprunck S. Manual isolation of living cells from the Arabidopsis thaliana female gametophyte by micromanipulation. *Methods Mol Biol.* 1669;2017:221–34.
57. Frawley LE, Orr-Weaver TL. Polyploidy. *Curr Biol.* 2015;25:R353–8.
58. Dresselhaus T, Franklin-Tong N. Male-female crosstalk during pollen germination, tube growth and guidance, and double fertilization. *Mol Plant.* 2013;6:1018–36.
59. Yu HJ, Hogan P, Sundaresan V. Analysis of the female gametophyte transcriptome of Arabidopsis by comparative expression profiling. *Plant Physiol.* 2005;139:1853–69.
60. Sprunck S, Baumann U, Edwards K, Langridge P, Dresselhaus T. The transcript composition of egg cells changes significantly following fertilization in wheat (*Triticum aestivum* L.). *Plant J.* 2005;41:660–72.
61. Ishikawa R, Ohnishi T, Kinoshita Y, Eiguchi M, Kurata N, Kinoshita T. Rice interspecies hybrids show precocious or delayed developmental transitions in the endosperm without change to the rate of syncytial nuclear division. *Plant J.* 2011;65:798–806.
62. Chen J, Strieder N, Krohn NG, Cypris P, Sprunck S, Engelmann JC, Dresselhaus T. Zygotic genome activation occurs shortly after fertilization in maize. *Plant Cell.* 2017;29:2106–25.
63. Wuest SE, Vijverberg K, Schmidt A, Weiss M, Gheyselinck J, Lohr M, Wellmer F, Rahnenfuhrer J, von Mering C, Grossniklaus U. Arabidopsis female gametophyte gene expression map reveals similarities between plant and animal gametes. *Curr Biol.* 2010;20:506–12.
64. Schmid MW, Schmidt A, Klostermeier UC, Barann M, Rosenstiel P, Grossniklaus U. A powerful method for transcriptional profiling of specific cell types in eukaryotes: laser-assisted microdissection and RNA sequencing. *PLoS One.* 2012;7:e29685.
65. Schmidt A, Schmid MW, Klostermeier UC, Qi W, Guthorl D, Sailer C, Waller M, Rosenstiel P, Grossniklaus U. Apomictic and sexual germline development differ with respect to cell cycle, transcriptional, hormonal and epigenetic regulation. *PLoS Genet.* 2014;10:e1004476.
66. Schon MA, Nodine M. Widespread contamination of Arabidopsis embryo and endosperm transcriptome datasets. *Plant Cell.* 2017;29:608–17.
67. Huang Q, Dresselhaus T, Gu H, Qu LJ. Active role of small peptides in Arabidopsis reproduction: expression evidence. *J Integr Plant Biol.* 2015;57:518–21.
68. Marshall E, Costa LM, Gutierrez-Marcos J. Cysteine-rich peptides (CRPs) mediate diverse aspects of cell-cell communication in plant reproduction and development. *J Exp Bot.* 2011;62:1677–86.
69. Brady SM, Orlando DA, Lee JY, Wang JY, Koch J, Dinneny JR, Mace D, Ohler U, Benfey PN. A high-resolution root spatiotemporal map reveals dominant expression patterns. *Science.* 2007;318:801–6.
70. Berger F, Hamamura Y, Ingouff M, Higashiyama T. Double fertilization - caught in the act. *Trends Plant Sci.* 2008;13:437–43.
71. Tian HQ, Yuan T, Russell SD. Relationship between double fertilization and the cell cycle in male and female gametes of tobacco. *Sex Plant Reprod.* 2005;17:243–52.
72. Zhurinsky J, Leonhard K, Watt S, Marguerat S, Bahler J, Nurse P. A coordinated global control over cellular transcription. *Curr Biol.* 2010;20:2010–5.
73. Sundaresan V, Alandete-Saez M. Pattern formation in miniature: the female gametophyte of flowering plants. *Development.* 2010;137:179–89.
74. Yang WC, Shi DQ, Chen YH. Female gametophyte development in flowering plants. *Annu Rev Plant Biol.* 2010;61:89–108.
75. Xiong Y, Sheen J. The role of target of rapamycin signaling networks in plant growth and metabolism. *Plant Physiol.* 2014;164:499–512.
76. Brink RA, Cooper DC. Double fertilization and development of the seed in angiosperms. *Bot Gaz.* 1940;102:1–25.
77. Ikeda Y, Kinoshita Y, Susaki D, Ikeda Y, Iwano M, Takayama S, Higashiyama T, Kakutani T, Kinoshita T. HMG domain containing SSRP1 is required for DNA demethylation and genomic imprinting in Arabidopsis. *Dev Cell.* 2011;21:589–96.
78. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics.* 2013;29:15–21.
79. Smith T, Heger A, Sudbery I. UMI-tools: modeling sequencing errors in unique molecular identifiers to improve quantification accuracy. *Genome Res.* 2017;27:491–9.
80. Huang da W, Sherman BT, Lempicki RA. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res.* 2009;37:1–13.
81. Kim D, Paggi JM, Park C, Bennett C, Salzberg SL. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat Biotechnol.* 2019;37:907–15.
82. Pertea M, Pertea GM, Antonescu CM, Chang TC, Mendell JT, Salzberg SL. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat Biotechnol.* 2015;33:290–5.
83. Clough SJ, Bent AF. Floral dip: a simplified method for *Agrobacterium*-mediated transformation of *Arabidopsis thaliana*. *Plant J.* 1998;16:735–43.
84. Song Q, Ando A, Jiang J, Ikeda Y, Chen ZJ. Single-cell RNA-seq analysis reveals ploidy-dependent and cell-specific transcriptome changes in Arabidopsis female gametophytes. vol. SRP160651; NCBI SRA: <https://trace.ncbi.nlm.nih.gov/Traces/sra/?study=SRP160651>; 2020.
85. Song Q, Ando A, Jiang J, Ikeda Y, Chen ZJ. Single-cell RNA-seq analysis reveals ploidy-dependent and cell-specific transcriptome changes in Arabidopsis female gametophytes. GitHub: https://github.com/qingxinsong/single_cell; 2020.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.