the plant journal

SEB Society for Experimental Biology

The Plant Journal (2018) doi: 10.1111/tpj.14041

Transcriptomics analyses reveal the molecular roadmap and long non-coding RNA landscape of sperm cell lineage development

Lingtong Liu[1,†], Yunlong Lu[1,2,†], Liqin Wei[1,†], Hua Yu[2,3,†], Yinghao Cao[3], Yan Li[3], Ning Yang[1], Yunyun Song[1,2], Chengzhi Liang[3,*] and Tai Wang[1,2,*]

[1]Key Laboratory of Plant Molecular Physiology, Institute of Botany, Chinese Academy of Sciences, Beijing 100093, China,
[2]College of Life Science, University of Chinese Academy of Sciences, Beijing 100049, China, and
[3]Research Center for Plant Genomics, Institute of Genetics and Developmental Biology, Chinese Academy of Sciences, Beijing 100101, China

Received 26 March 2018; accepted 19 July 2018.
*For correspondence (e-mail twang@ibcas.ac.cn; cliang@genetics.ac.cn).
†These authors contributed equally to this work

SUMMARY

Sperm cell (SC) lineage development from the haploid microspore to SCs represents a unique biological process in which the microspore generates a larger vegetative cell (VC) and a smaller generative cell (GC) enclosed in the VC, then the GC further develops to functionally specified SCs in the VC for double fertilization. Understanding the mechanisms of SC lineage development remains a critical goal in plant biology. We isolated individual cells of the three cell types, and characterized the genome-wide atlas of long non-coding (lnc) RNAs and mRNAs of haploid SC lineage cells. Sperm cell lineage development involves global repression of genes for pluripotency, somatic development and metabolism following asymmetric microspore division and coordinated upregulation of GC/SC preferential genes. This process is accompanied by progressive loss of the active marks H3K4me3 and H3K9ac, and accumulation of the repressive methylation mark H3K9. The SC lineage has a higher ratio of lncRNAs to mRNAs and preferentially expresses a larger percentage of lncRNAs than does the non-SC lineage. A co-expression network showed that the largest set of lncRNAs in these nodes, with more than 100 links, are GC-preferential, and a small proportion of lncRNAs co-express with their neighboring genes. Single molecular fluorescence *in situ* hybridization showed that several candidate genes may be markers distinguishing the three cell types of the SC lineage. Our findings reveal the molecular programming and potential roles of lncRNAs in SC lineage development.

Keywords: sperm cell lineage, cell commitment, cell fate, molecular roadmap, long non-coding RNAs. *Solanum lycopersicum*.

INTRODUCTION

In animals, male germ cells are determined early in embryogenesis and remain a distinct population of stem cells for the production of sperm throughout reproductive life (Magnusdottir and Surani, 2014). In contrast, flowering plants do not have a distinct germ line during embryogenesis and the vegetative growth phase; they produce sperm cells (SCs) in anthers only as the plant develops into the reproductive phase (Berger and Twell, 2011). In this phase, the haploid microspore generated by meiosis of pollen mother cells is considered a pluripotent cell and precursor of the generative cell (GC) and SCs (Berger and Twell, 2011). Microspore asymmetric division is determinative and immediately confers the differential identity and fate of the two daughter cells. The larger vegetative cell (VC) is terminally differentiated, and functions as a helper cell in supporting gamete development and transporting SCs into the embryo sac for double fertilization. The smaller GC is engulfed in the VC, and further divides mitotically to produce two SCs (Berger and Twell, 2011). The series of cell populations from microspores to GCs and finally to SCs represents a well-characterized SC lineage. Sperm cell lineage development involves fine-tuned cell commitment, establishment of cell identity and fate and SC specification, which are basic biological issues.

Transcriptomic studies of pollen from diverse plant species including Arabidopsis and rice have characterized

gene expression profiles for pollen development (Rutley and Twell, 2015). The transcriptome is significantly reduced in complexity as pollen develops, with an increased number of stage-preferential transcripts (Honys and Twell, 2004; Borges *et al.*, 2008; Wei *et al.*, 2010; Rutley and Twell, 2015). The mature pollen is enriched in mRNAs involved in cell wall modification, the cytoskeleton, defense and stress responses, in accordance with the requirements of pollen tube growth and interaction with pistil cells. Because the GC and SCs are engulfed by the VC, the dissection of gene expression profiles in the two cell types is challenging. Several studies have reported expressed sequence tags of SCs from maize, *Plumbago* and tobacco (Engel *et al.*, 2003; Gou *et al.*, 2009; Xin *et al.*, 2011), and GCs from lily (Okada *et al.*, 2006). The genome-wide profiling of SCs from Arabidopsis and rice realized by isolating SCs provided a transcriptomic insight into SC specification (Borges *et al.*, 2008; Slotkin *et al.*, 2009; Russell *et al.*, 2012; Anderson *et al.*, 2013). These studies also showed transcriptional repression of somatic and metabolic genes in SCs and mature pollen compared with the microspore (Rutley and Twell, 2015). Genetic studies of Arabidopsis have characterized several genes/proteins involved in male germ cell development. The lateral organ boundary domain (LBD) protein SCP, protein kinase TIO, kinesin-12A/B, GEM1 and GEM2 are important for asymmetric division of the microspore and cell fate (Park *et al.*, 1998, 2004; Oh *et al.*, 2005, 2010; Lee *et al.*, 2007). CDKA;1 and F-box protein FBL17 are responsible for division of the GC (Iwakawa *et al.*, 2006; Kim *et al.*, 2008), whereas the transcription factors (TFs) DUO1 and DUO3 are responsible for both division of the GC and SC specification (Brownfield *et al.*, 2009a,b; Borg *et al.*, 2011). Pollen deficient in CDKA;1, FBL17, DUO1, DUO3 or DAZ1/DAZ2 has only one germ cell. The single germ cell in pollen deficient in CDKA;1 or FBL17 can fertilize the egg cell, but germ cells deficient in DUO1, DUO3 or DAZ1/DAZ2 cannot (Iwakawa *et al.*, 2006; Nowack *et al.*, 2006; Kim *et al.*, 2008; Kasahara *et al.*, 2012; Borg *et al.*, 2014). Clarifying the differences between the GC and SCs requires cell type-specific markers. However, the difficulty in obtaining all three cell types of the SC lineage from the same species has constrained our ability to characterize molecular markers of GCs and SCs and the molecular program for SC lineage development.

Long non-coding RNAs (lncRNAs) are a major type of regulatory RNA that has important roles in diverse biological processes in mammals and plants (Quinn and Chang, 2016). The mammalian lncRNAs Xist, H19 and HOTAIR function in the precise control of embryogenesis and cell proliferation (Penny *et al.*, 1996; Gabory *et al.*, 2010; Gupta *et al.*, 2010). At the genome-wide level, approximately one-third of human lncRNAs are specifically expressed in testes (Cabili *et al.*, 2011). Plant lncRNAs such as COLDAIR and LDMAR are involved in the epigenetic regulation of flowering and male fertility (Heo and Sung, 2011; Ding *et al.*, 2012). Many rice lncRNAs are significantly enriched in anthers (Zhang *et al.*, 2014). These data suggest that lncRNAs may have highly specialized roles in the differentiation of specific cell types, but these roles are largely unknown.

Here, we isolated tomato microspores, GCs and SCs for transcriptomic analysis. In total, 31 931 lncRNAs and 38 980 mRNAs were expressed in SC lineage cells and four other cells/tissues. The locations of lncRNA-coding genes were highly physically correlated with protein-coding genes. Transcription reprogramming during SC lineage development involved the progressive elimination of active histone marks and the accumulation of repressive histone marks. Our data support that global repression of somatic and metabolic genes is a significant molecular signature for germ cell commitment and gamete development across plants and mammals. The co-expression network of lncRNAs and mRNAs provides clues for deciphering the functions of lncRNAs expressed in SC lineage cells. Finally, by using these data we identified molecular markers distinguishing each cell type of the SC lineage.

RESULTS

Long non-coding RNA and mRNA analysis of the SC lineage

Transcriptome profiling of the SC lineage requires an appropriate experimental system that permits isolation of microspores, GCs and SCs from the same plant species. Mature tomato pollen is bicellular at anthesis, which enables isolation of GCs from mature pollen and SCs from cultured pollen tubes where SCs are generated. We have reported efficient methods for isolating GCs and SCs in tomato (Lu *et al.*, 2015). Here, we prepared high-purity SC lineage cells for transcriptomic study (Figure 1a). In addition, four different control materials were used: callus (pluripotent cell population), fully expanded leaves (differentiated somatic cells), pistil (highly heterogeneous tissue composed of female gametes and differentiated somatic cells) and mature pollen grain (MPG). The MPG is significantly represented by the VC which supplies most of the bulk and cytoplasm, and is thus generally considered to reflect the main molecular characteristics of the VC (Xu *et al.*, 1999).

Genome-wide strand-specific RNA sequencing (ssRNA-seq) of all seven samples, with three independent biological replicates each, resulted in more than 1.3×10^8 clean reads for each replicate (Table S1 in the online Supporting Information). To comprehensively characterize the lncRNA and mRNA programming underlying SC lineage development, we used the 21 ssRNA-seq datasets and 88 publicly available tomato RNA-seq datasets to re-annotate the

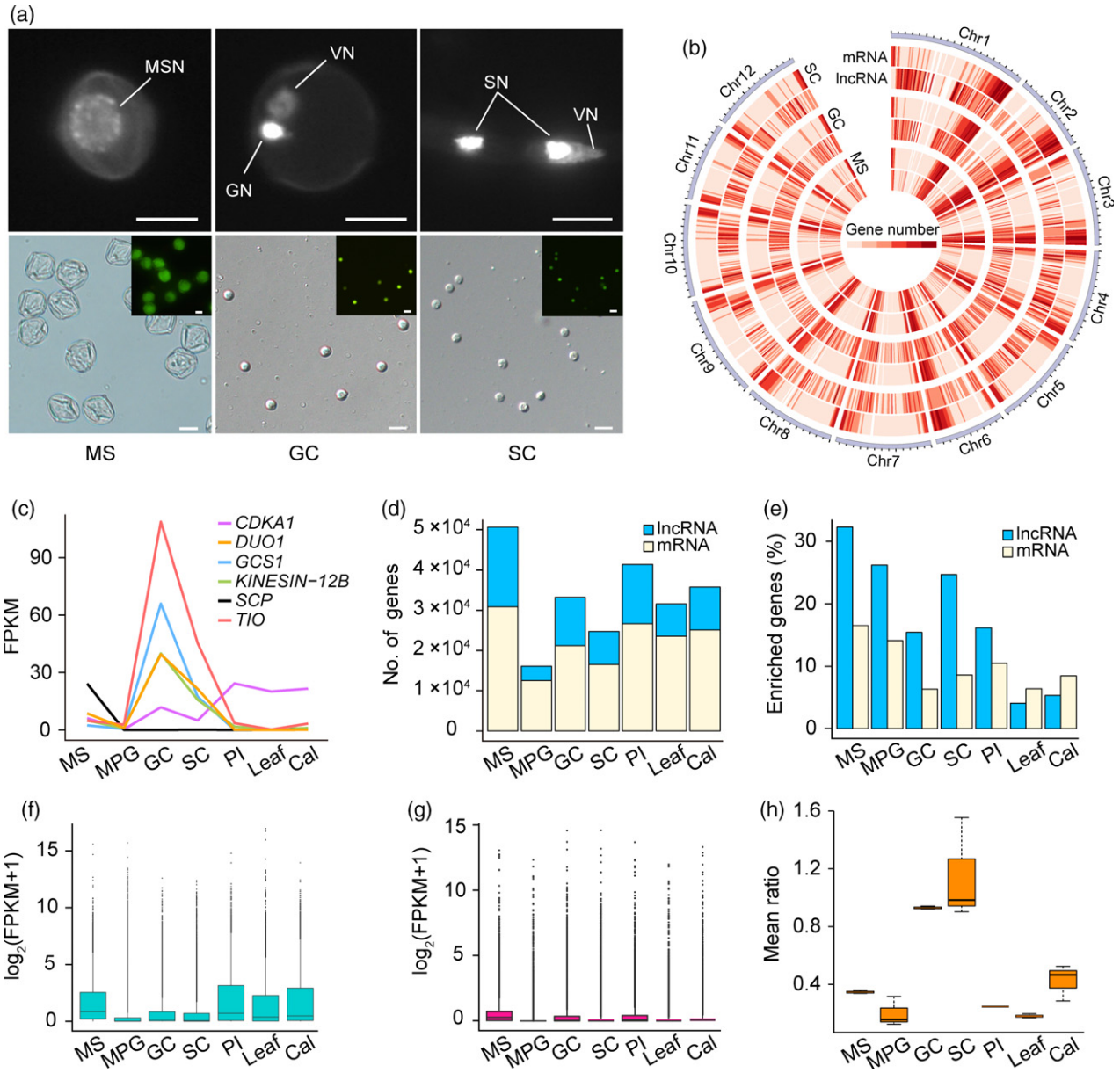


Figure 1. Genome-wide identification of mRNAs and long non-coding RNAs (lncRNAs) in the sperm cell (SC) lineage.
(a) Upper panel: 4,6-diamidino-2-phenylindole staining of the microspore (MS), mature pollen grain (MPG) and pollen tube to visualize the nucleus. MSN, microspore nucleus. VN, vegetative cell nucleus. GN, generative cell nucleus, SN, sperm cell nucleus. Lower panel: purified MS, generative cell (GC), and SC. Insets indicate the viability of MS, GC and SC tested by fluorescein diacetate staining. Scale bars, 10 μm. (b) Circos plot showing the number of expressed mRNAs (outer circles) or lncRNAs (inner circles) in the genome of the SC lineage. Transcriptional hotspots are indicated in red. (c) Expression patterns of six representative functionally known genes for SC lineage development. (d) Number of mRNAs and lncRNAs expressed in indicated cells and tissues. PI, pistil; Cal, callus. (e) The percentage of cell/tissue-enriched mRNAs and lncRNAs in corresponding mRNomes and lncRNomes. (f, g) Box-plot representation of mRNA (f) and lncRNA (g) abundance in indicated cells and tissues. (h) Mean ratio of lncRNA to mRNA expression in indicated cells and tissues.

tomato genome. The re-annotation generated 53 666 protein-coding and 53 651 lncRNA-coding genes. As a comparison, the re-annotation covered 25 104 (72.3%) genes in the tomato genome annotation release (Tomato Genome, 2012) with many extra genes not present in ITAG2.4. We added the extra re-annotated genes to ITAG2.4 to form a complete gene set and computed the expression level of each gene. We detected the expression of 38 980 protein-coding and 31 931 lncRNA-coding genes in these samples [fragments per kilobase of exon per million reads (FPKM) > 0.09 in all three biological replicates of at least one sample] (Table S2). A total of 43.6% protein-coding genes and 30.4% lncRNA genes were in a 6-Mb region of chromosome ends, with a correlation coefficient 0.85

(Figures 1b and S1a,b), indicating a high regional correlation between expressed mRNA and lncRNA genes. Among these lncRNAs, 90.47% were intergenic, with the remainder classified as sense, antisense or intronic lncRNAs (Figure S1c). The lncRNAs were shorter than mRNAs, with fewer but longer exons (Figure S1d). We analyzed expression profiles of transcripts in SC lineage cells by comparing the relative abundance of the transcripts (no spike-in). We verified the expression of these identified transcripts by using quantitative (q)RT-PCR with a random selection of 29 genes (11 lncRNAs, 18 mRNAs); 83% showed consistent expression profiles as revealed by RNA sequencing (RNA-seq) (Figure S1e, Table S3). Hence, the RNA-seq and genome re-annotation were faithful. Furthermore, the expression profiles of functionally known genes for SC lineage development, including *SCP*, *DUO1*, *GCS1* and *TIO* (Figure 1c), were as previously reported (Oh *et al.*, 2005, 2010; Mori *et al.*, 2006; Borg *et al.*, 2011). Thus, the experimental designs allowed us to dissect the gene expression patterns for SC lineage development and molecular signatures associated with each cell type of the SC lineage.

Long non-coding RNAs are abundantly expressed in individual cells of the SC lineage

Microspores expressed the highest diversity of mRNAs and lncRNAs: 30 906 mRNAs and 19 709 lncRNAs (hereafter, all expressed mRNAs and lncRNAs in a cell/tissue are called mRNome and lncRNome, respectively). The diversity of the mRNome and lncRNome was substantially reduced from the microspore to the GC and ultimately SCs (16 561 mRNAs and 8204 lncRNAs), with the lowest diversity in the MPG (Figure 1d). The proportion of cell-type-enriched lncRNAs was significantly greater in the lncRNome of the SC lineage and the MPG and pistil than in the corresponding somatic cell mRNome (Figure 1e, Table S4).

The downregulation of transcriptome diversity also coincided with a decrease in transcript abundance. The microspore, along with leaf, callus and pistil, expressed high levels of transcripts, which were significantly reduced in the GC and SCs, with the lowest levels in the MPG (Figure 1e,g). Levels were lower for lncRNAs than mRNAs in all samples (Figure 1e,g), but lncRNAs displayed greater cell/tissue-type specificity than mRNAs (Figure S1f). Moreover, the mean ratio of lncRNAs to mRNAs (mean FPKM of lncRNAs to mRNAs) was greater in the SC lineage than MPG, pistil, leaf and callus, and was increased from 0.35 in the microspore to 0.93 in the GC and finally 1.15 in SCs (Figure 1h). So lncRNAs were abundantly expressed in the SC lineage, especially in the GC and SCs.

The microspore retains the molecular features of somatic cells

A microspore is cytologically considered a type of pluripotent cell that develops into an embryo-like structure upon stress or hormone induction (Honys and Twell, 2004; Segui-Simarro and Nuez, 2007; Daghma *et al.*, 2014; Seifert *et al.*, 2016); however, molecular evidence supporting this concept is still limited. Principal component analysis (PCA) and correlation matrix analysis showed that the mRNome of the microspore is closer to that of callus ($r = 0.71$), leaf ($r = 0.65$) and pistil ($r = 0.72$) than that of the GC ($r = 0.46$), SCs ($r = 0.35$) and MPG ($r = 0.40$) (Figures 2a and S2a, Table S5). Furthermore, the microspore shared a set of 5995 mRNAs with callus, leaf and pistil, with significantly higher expression than in the GC, SCs and MPG (Figure 2b). The enriched Gene Ontology (GO) terms included post-embryonic development, multicellular organismal development and diverse metabolism (Figure 2c), which suggests highly active somatic or 'vegetative' biological pathways in the microspore.

The microspore shared 22 656 mRNAs with callus, representing 73.3% and 90.1% of the respective mRNomes. This dataset includes marker genes for shoot stem cells (*WUS*, *STM* and *KNAT6*), root stem cells (*WOX5*) and cambium stem cells (*WOX4*), which are master regulators for maintaining the pluripotent state of respective stem cell types (Greb and Lohmann, 2016). Almost all these genes were preferentially expressed in the microspore or callus but significantly downregulated in subsequent SC lineage cells (Figure 2d). Moreover, recent studies have demonstrated LBD proteins to be important transcription factors (TFs) for plant regeneration and cell differentiation (Xu *et al.*, 2016). Among 41 *LBD*s (e.g. *AS2*, *LOB*, *SCP*) annotated in the tomato genome, 32 and 31 were expressed in the microspore and callus, respectively, with only 14 and 9 expressed in the GC and SCs, respectively (Figure S2b). The mRNAs co-expressed with these microspore- or callus-preferential LBDs shared GO terms related to organ development, response to stimuli and metabolism, but the microspore showed the unique GO terms of reproduction, pollen development and transcription (Figure S2c,d), which indicates an LBD-related network for regulation of gene transcription in the microspore. These data suggest that the microspore retains several pluripotent regulators but acquires a state for shifting to male germ development.

Generally, SC lineage development is reminiscent of stomatal differentiation, whereby a meristemoid divides asymmetrically to produce a pavement cell and a guard mother cell, which divides symmetrically to produce two guard cells (Lau and Bergmann, 2012). However, unlike the meristemoid, which can undergo several rounds of asymmetrical division, a microspore divides asymmetrically only once. We found that the microspore expressed *MUTE*, *SPCH* and *FAMA* (*SlbHLH146*), three pivotal TFs required for stomatal lineage development in Arabidopsis (Lau and Bergmann, 2012). *MUTE*, which promotes cell differentiation and termination of asymmetrical division, was still expressed in the GC and SCs; however, expression of

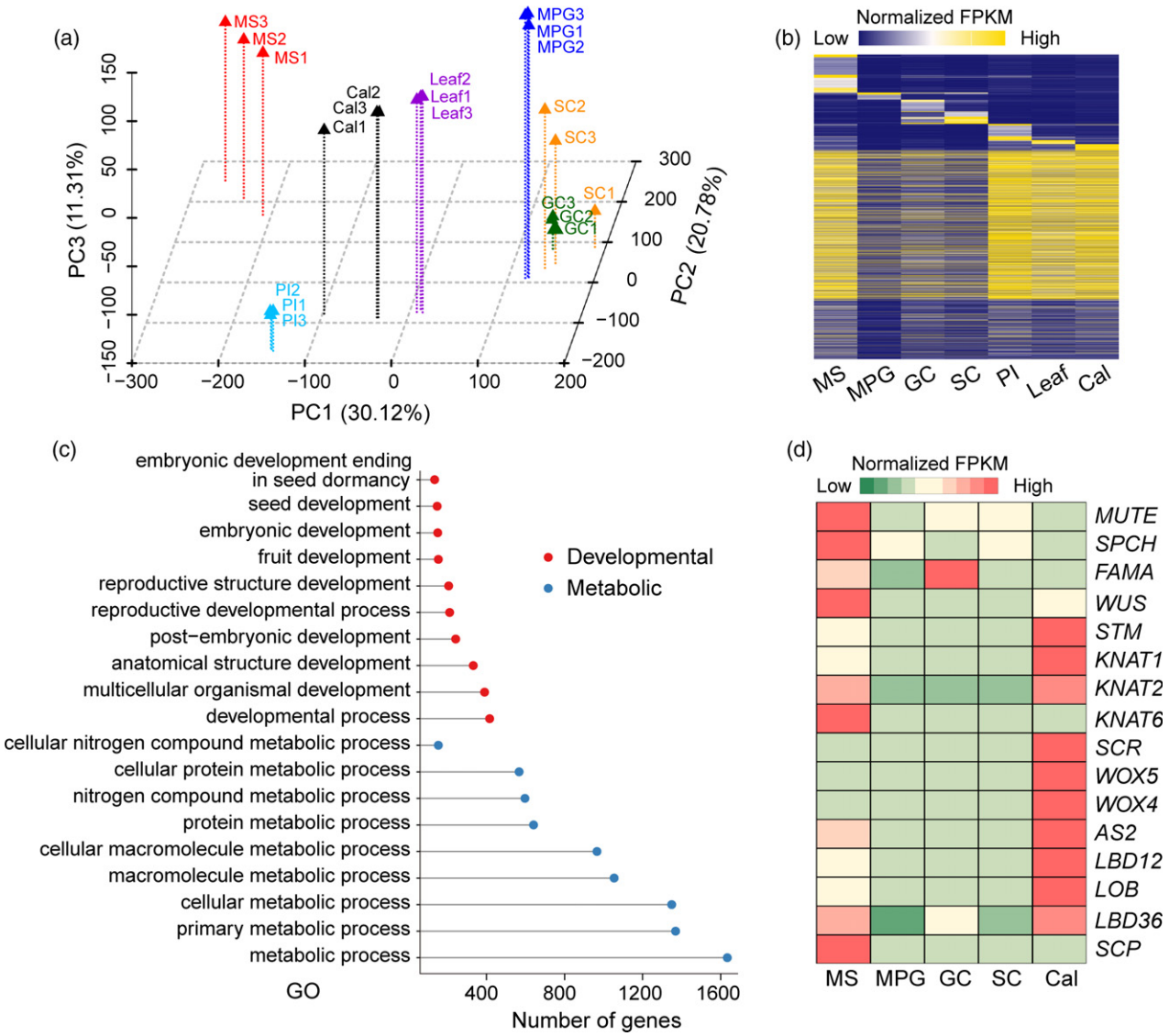


Figure 2. The microspore retains molecular features of somatic cells.
(a) Principal component analysis of global gene expression in indicated cells and tissues, each with three biological replicates.
(b) Heatmap representation of mRNA expression across the indicated cells and tissues. The microspore (MS) shares a large set of mRNAs with pistil, leaf and callus as compared with the generative cell (GC), sperm cell (SC) and mature pollen grain (MPG). PI, pistil; Cal, callus.
(c) Gene Ontology (GO) categories ($P < 0.001$, false discovery rate < 0.001) of mRNAs preferentially expressed [>five-fold increase in expression, fragments per kilobase of exon per million reads (FPKM) > 0.5] across MS, PI, leaf and Cal. The top GO terms are shown.
(d) Expression dynamics of representative genes with Arabidopsis homologs that function in cell differentiation. All genes are denoted by the names of the corresponding Arabidopsis homologs.

SPCH, which drives asymmetric division, was barely detectable in the GC. *FAMA*, which controls differentiation of guard mother cells, was expressed in the GC but not SCs (Figure 2d). The expression patterns of the three genes in the SC lineage may explain why microspores have limited proliferation ability.

Reprogramming following asymmetric division

After asymmetric division of the microspore, a GC further develops into two SCs. A key question is how gene expression contributes to commitment to the SC lineage. During this process, most TFs expressed at a high level in the microspore were sharply downregulated in the GC and remained at low levels in SCs (Figure S3a). Concomitantly downregulated mRNAs were associated with signaling of almost all phytohormones including auxin, cytokinin, gibberellic acid, brassinosteroid, abscisic acid, ethylene, salicylic acid and jasmonate (Figure S3b).

We focused on the common and different characteristics of mRNomes of the microspore, GC and SCs. In total,

12 492 of 35 396 mRNAs detected in the three cell types were constantly expressed across the three cell types, and showed enrichment in the GO terms of cell cycle, DNA metabolic process and cell differentiation (Figures 3a and S4), which suggests that these pathways are a common theme in three cell types. Consistent with the microspore having molecular features of somatic cells (see above), mRNAs expressed only in the microspore were associated with GO terms associated with somatic cells and vegetative tissue. However, mRNAs commonly expressed only in the GC and SCs were enriched in the GO terms anatomical structure development and cellular component morphogenesis; mRNAs expressed only in the GC or SCs had few enriched GO terms, possibly because of the lack of available functional studies (Figures 3a and S4). These results indicate that distinct mRNA sets have markedly different functional features.

Next, we examined changes in mRNA expression patterns in progressive cell pairs of the SC lineage. A total of 9244 distinct mRNAs were downregulated (greater than four-fold change in expression, counts per million (cpm) > 1, $P < 0.05$), with 5584 upregulated in the GC versus the microspore (Figure 3b); similarly, we observed sharply downregulated mRNAs in the MPG versus the microspore (Figure S5). From the GC to SCs, 6230 mRNAs were downregulated and only 1582 upregulated (Figure 3c). The progressively downregulated or upregulated genes in the GC versus the microspore and SCs versus the GC did not overlap significantly (Figure 3d). These downregulated genes are enriched in vegetative tissue/organ development, hormone signaling and metabolism. Consistently, the hierarchical clustering of expressed mRNAs in the SC lineage revealed the largest set of mRNAs ($n = 25\,438$, clusters 1, 2 and 3) that were expressed at high levels in the microspore and decreased to low levels in the GC and SCs. The genes featured significantly enriched GO terms for shoot morphogenesis, meristem maintenance, metabolism, response to hormone and stress, protein localization, telomere maintenance and transcription (including 668 TFs) (Figure 3e). Only a few genes were expressed at similar levels ($n = 1845$, cluster 4) in the microspore and GC. Genes in clusters 5, 6, 7 and 8 showed increased expression after asymmetric division, featuring GO terms for histone H3K9 methylation, DNA replication, gene silencing, cell wall dynamics, transporter activity and pollen tube growth. Genes with the lowest expression (clusters 9 and 10) in the GC featured GO terms for protein modification, transmembrane transport and signal transduction. However, genes showing the highest expression in the GC (clusters 11 and 12) featured GO terms for cell cycle, regulation of cellular process and stress response (Figure 3e). The GO terms enriched in the SC lineage differed from those in the MPG, which largely involved polar pollen tube growth and cell communication, indicating that the MPG reflects the features of the VC to some extent (Figure S5), and thereby implying the involvement of reprogramming in mRNome differentiation of the SC lineage. As for genes associated with the GO term cell cycle, 348, 332 and 279 were significantly downregulated in pollen compared with the microspore, GC and SCs, respectively. No cell cycle genes were significantly upregulated in pollen. This finding is consistent with the notion that the VC exits the cell cycle after asymmetric division. Furthermore, we evaluated conservation and diversity of the tomato SC mRNome and other SC mRNomes from rice and Arabidopsis (Borges *et al.*, 2008; Anderson *et al.*, 2013). Although these mRNomes were divergent, the pairwise pairs had conserved mRNA sets (Figure S6a). The GO terms cell cycle, ubiquitin pathway, DNA modification and repair, and chromosome organization were enriched in the conserved mRNA sets, which is consistent with the observation that mRNAs for these terms were enriched in SCs of rice and Arabidopsis (Borges *et al.*, 2008; Russell *et al.*, 2012; Anderson *et al.*, 2013). Conserved mRNA sets were significantly enriched in GO categories involved in negative regulation of transcription, gene expression, metabolism and biological process, and diverse epigenetic pathways, including H3K9 methylation, RNA-mediated gene silencing and chromatin silencing (Figure S6b), which suggests that these conserved pathways are critical for SC specification.

Chromatin-related factors involved in progressive transcriptional restriction

To gain insight into global transcriptional repression, we investigated the expression of five DNA-dependent RNA polymerases (Pols I–V) and chromatin-related factors in the SC lineage (Figure 4). Pols I–V were hardly detectable in the MPG, but were found at high levels in the microspore, with high levels of Pols II, III and V in other tissues (Figure 4a). The expression of Pols I and III, which mainly govern rRNA and tRNA synthesis (Paule and White, 2000), was barely detectable in the GC and SCs. The expression of Pol II, implicated in mRNA and lncRNA synthesis, was detectable in the GC and became undetectable in SCs. Pols IV and V, specialized for production of non-coding RNAs involved in DNA methylation and gene silencing (Ream *et al.*, 2009; Haag and Pikaard, 2011), was low in the GC then upregulated in SCs (Figure 4a). In addition, the GC and SCs showed weak expression of 25S rRNA, which was highly abundant in the microspore and other cells and tissues (Figure 4a). We further examined the protein levels of Pol II and the 40S ribosomal protein S14-1 and found that expression of Pol II and S14-1 changed from high in the microspore to low in the GC, and undetectable in SCs. The two proteins were found at high levels in the MPG (Figure 4b). The difference in expression pattern of Pol II mRNAs and proteins in the MPG is consistent with the

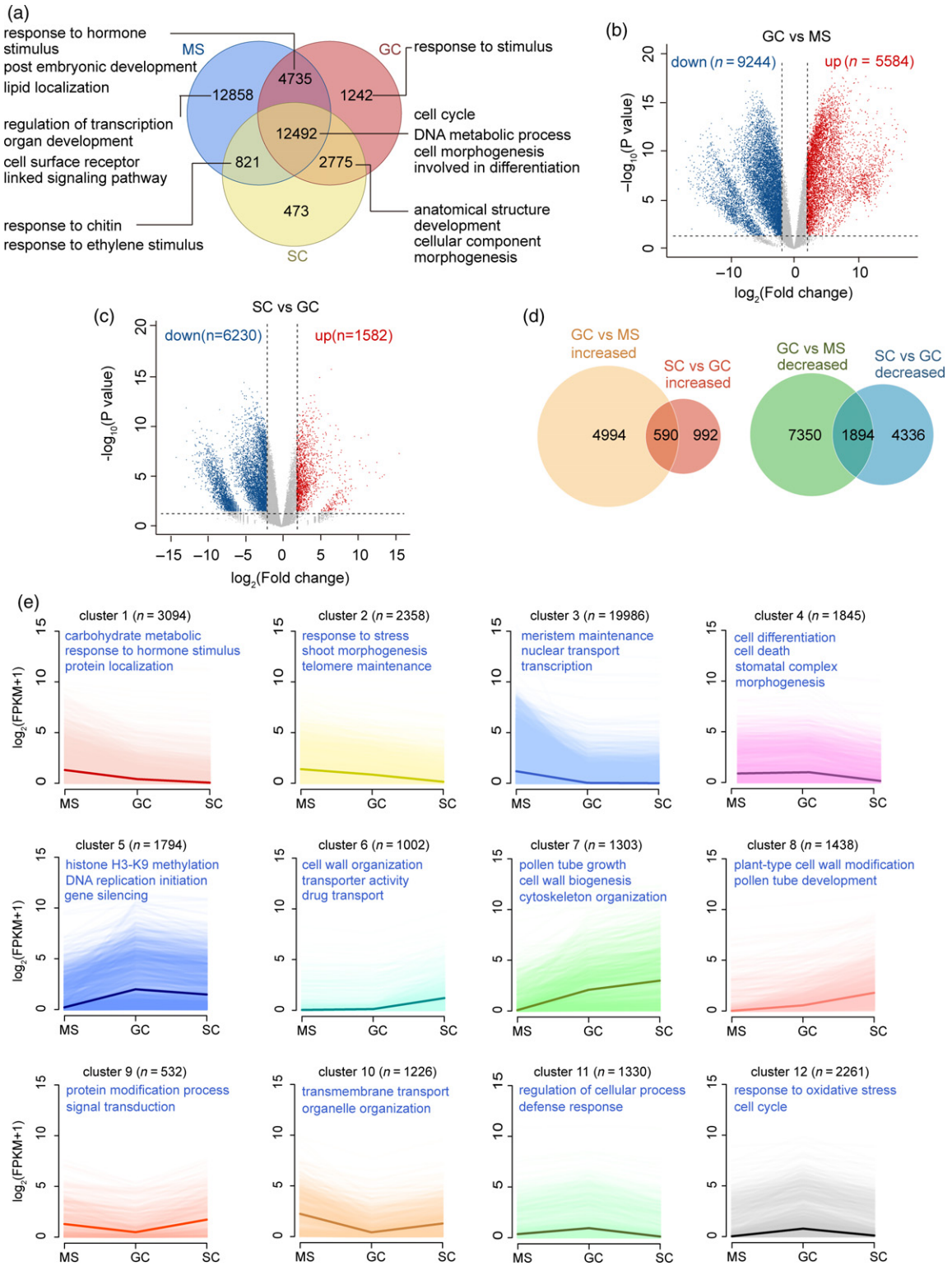


Figure 3. Dynamics of mRNA expressed in the sperm cell (SC) lineage.
(a) Pair-wise comparison of mRNA profiles of the three cell types. The Venn diagram shows the association of the three mRNA profiles and representative top Gene Ontology (GO) terms ($P < 0.01$, false discovery rate < 0.05).
(b, c) Number of differentially expressed mRNAs (P-value < 0.01, greater than four-fold change in expression) in progressive pairs of the three cell types.
(d) Association between progressively up- or downregulated mRNAs [generative cell (GC) versus microspore (MS) and SC versus GC].
(e) Gene expression patterns clustered by expression changes in SC lineage. Representative GO terms are indicated in blue.

previous notion that pollen germination and early tube growth is largely independent of transcription but dependent on protein synthesis (Honys and Twell, 2004). Collectively, these results suggest progressive inhibition of transcription and translation from the microspore to GC and SCs and an important role of RNA-directed DNA methylation (RdDM) and gene silencing in this process.

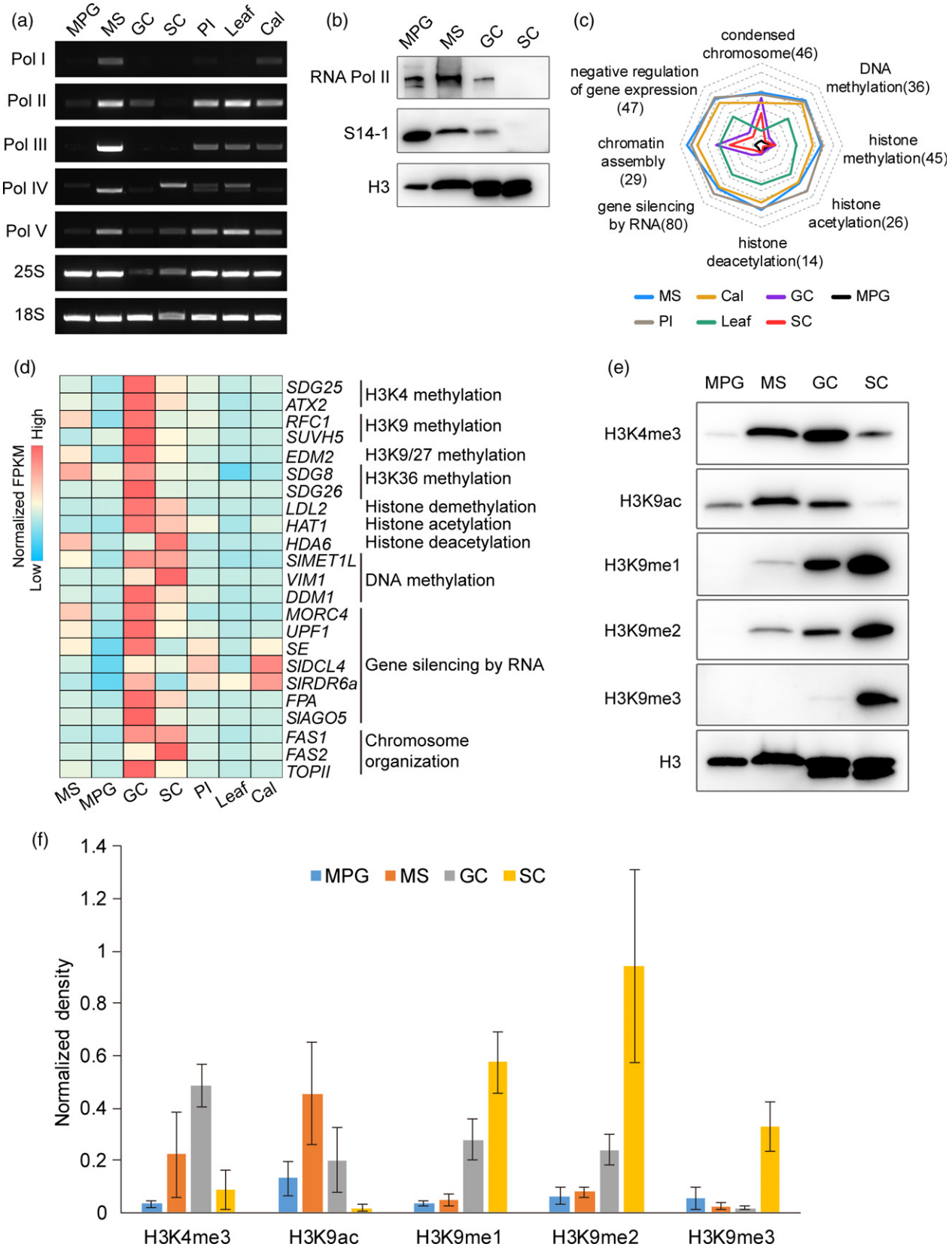
Generally, the expression of genes for different chromatin-related pathways was at high levels in the microspore and callus and pistil but slightly decreased (for chromatin assembly and condensation) or sharply decreased (for other pathways) in the GC and SCs, with the lowest levels in the MPG (Figure 4c). These genes were involved in histone acetylation and deacetylation (e.g. *TAF9*, *HDA1*, *HDA9*), small RNA-mediated gene silencing (e.g. *SlAGO1b*, *SlAGO4b*, *SAD2*, *SlDEAH7*), DNA and histone methylation (e.g. *SlDML1*, *JMJ22*, *LSD1*) and chromatin assembly and condensation (Figure 4c). This concomitant change was also observed for the polycomb group genes *SWN* (*SlEZ1*), *CLF* (*SlEZ2)*, *EMF2* and *MSI1*, which silence target genes during gametophytic or sporophytic development (Hennig and Derkacheva, 2009); and multiple H2A variants including *HTA3*, *HTA5*, *HTA7*, *HTA9* and *HTA10*, the spatial–temporal expression of which represents an important mechanism regulating gene expression (Skene and Henikoff, 2013).

Although genes for different chromatin-related pathways were generally downregulated or sharply downregulated after asymmetric division, a set of genes involved in *trans*-acting small interfering RNA (ta-siRNA)-mediated gene silencing, maintenance of the epigenetic state and chromatin organization were GC- and/or SC-preferential (Figure 4d). Among GC-preferential chromatin-related genes, *SE*, *SlDCL4* and *SlRDR6a* are required for accumulation of ta-siRNA or RdDM in plants (Yang *et al.*, 2006; Pumplin *et al.*, 2016). *UPF1* is involved in nonsense-mediated mRNA decay and RNA interference (Yoine *et al.*, 2006), *DDM1* in the maintenance of DNA and histone methylation (Jeddeloh *et al.*, 1998) and *SUVH5* in maintenance of H3K9 dimethylation and CMT3-mediated non-CG DNA methylation (Rajakumara *et al.*, 2011). Following their progressive development, SCs showed preferential expression of the chromatin assembly factors *FAS1* and *FAS2*, which function as histone chaperones to mediate the deposition of histones or histone variants in newly synthesized DNA (Kaya *et al.*, 2001), the methylation-dependent chromatin silencing regulator *VIM1* (Woo *et al.*, 2007) and the histone deacetylase gene *HDA6*, which is involved in large-scale gene silencing in nucleolar dominance (Earley *et al.*, 2006) (Figure 4d). We further examined the SC lineage for the dynamics of histone H3 modifications at lysines 4 and 9 (H3K4me3, H3K9ac, H3K9me1, H3K9me2 and H3K9me3), which are important histone marks required for gene activation or silencing (Liu *et al.*, 2010; Borg and Berger, 2015). Immunoblot analysis showed high expression of the active marks H3K4me3 and H3K9ac in the microspore and the GC but greatly decreased (H3K4me3) or barely detectable levels (H3k9ac) in SCs (Figure 4e,f). Expression of the repressive marks H3K9me1 and H3K9me2 was barely detectable in the microspore, increased sharply to high levels in the GC and peaked in SCs, whereas that of the repressive H3K9me3 was almost undetectable in the microspore and GC but was high in SCs (Figure 4e and f). Together with the observation that the complexity and transcript abundance of transcriptomes were progressively reduced from the microspore to the GC and SCs, these results suggest that these active and repressive histone marks are important for SC lineage development.

Defining candidate lncRNAs for SC lineage development

Hierarchical clustering classified differentially expressed lncRNAs into 12 expression patterns, which were generally similar to those of mRNAs (Figure S7a). Next, we evaluated differentially expressed lncRNAs in progressive pairs in the SC lineage (greater than four-fold in expression change, cpm > 1, $P < 0.01$). Impressively, lncRNAs showed preferentially upregulated and downregulated expression from the microspore to the GC, with relatively small sets of new lncRNA further upregulated and downregulated from the GC to SCs, respectively (Figure S7b and c). Clearly, lncRNAs showed a global transcriptional repression switch from the microspore to the GC and finally SCs. To investigate whether specific genomic regions were repressed or activated during SC lineage development, we calculated expressed gene diversity in regions of 1 Mb. A genomic region was defined as repressed or activated when gene diversity underwent more than a five-fold change in SCs compared with the GC or microspore. We found 15 regions to be repressed: six were in chromosome 9, with the largest decrease being 19-fold (Figure S8, Table S6). Gene

Figure 4. Global transcription restriction during sperm cell (SC) lineage development.
(a) The RT-PCR analysis of the largest subunits in five DNA-directed RNA polymerases(Pol) and 25S ribosomal RNA in indicated cells and tissues. 18S rRNA was used as an endogenous control. MPG, mature pollen grain; MS, microspore; GC, generative cell; PI, pistil; Cal, callus. (b) Immunoblot analysis of Pol II and the 40S ribosomal protein S14-1. Histone H3 was a loading control. Total protein extracted from the MPG (10 μg) and SC lineage cells (5 μg) were loaded in each lane and separated by 4–20% SDS-PAGE. (c) Radar chart showing the expression of genes related to epigenetic regulation. Eight Gene Ontology (GO) terms are arranged radially, and the median expression of genes associated with each GO term is depicted on the spoke. The outermost and innermost dashed lines indicate the highest and lowest expression, respectively. Numbers in brackets indicate the number of genes in each group. (d) Expression profiles of representative epigenetic regulators preferentially expressed in the GC and/or SCs. (e) Immunoblot analysis of histone H3 modification. Total protein extracted from the MPG (15 μg) and SC lineage cells (5 μg) was loaded in each lane. Representative results from three replicates are shown. (f) Quantification of Western blots with ImageJ. The integrated density of Western bands in (e) was normalized to that of H3. Data are means ± SD ($n = 3$).



 The Plant Journal, (2018), doi: 10.1111/tpj.14041

diversity was lower in the GC than in the microspore but higher than in SCs, which suggests progressive gene silencing in the repressed regions. In contrast, 17 regions were activated, and gene diversity in these regions was similar in the GC and SCs. Eight regions showed progressive activation in the microspore relative to the GC and SCs, whereas eight regions expressed the highest number of genes in the GC but not in SCs. Together, these results suggest immediate reprogramming following asymmetric division.

To gain insight into the functional relation between lncRNAs and mRNAs, we constructed their co-expression network by the law of 'guilt by association' (Rinn and Chang, 2012). This network consisted of 17 520 nodes and 705 897 links (Table S7). Among the co-expressed lncRNA and mRNA pairs, only 306 were located within a 100-kilobase neighboring region (Figure 5a). Then, we extracted 1129 lncRNAs with more than 100 links ('key lncRNAs') from the network. The largest portion (506) of these key lncRNAs were GC-preferential, the second largest (251) SC-preferential and the smallest (13) microspore-preferential (Figure S9a); the GC-preferential key lncRNAs were expressed at significantly higher levels than the other lncRNAs (Figure S9b). As an example, we visualized a subnetwork containing a set of key lncRNAs and GC- or SC-enriched mRNAs: the expression of lncRNAs was highly correlated with that of chromatin-related regulators and DUO1-targeted genes such as *GCS1*, *GEX2* and *DAA1* (Figure 5b).

Furthermore, we extracted mRNAs highly associated with cell type-enriched lncRNAs (confidence score > 0.7) from this network and revealed 1273 and 842 mRNAs that were co-expressed with GC- and SC-enriched lncRNAs, respectively. The 1273 GC-enriched lncRNA-associated mRNAs featured significantly enriched terms for cell cycle and chromosome organization. The 842 SC-enriched lncRNA-associated mRNAs featured enriched terms for cell cycle, chromosome organization and DNA metabolic process. Distinct from the SC lineage, the MPG showed 1534 enriched lncRNA-associated mRNAs, with the GO term pollen tube growth. As expected, leaf-enriched lncRNA-associated mRNAs were predominantly associated with photosynthesis. Callus-enriched lncRNA-associated mRNAs were mainly involved in response to hormone stimulus and regulation of transcription. Pistil-enriched lncRNA-associated mRNAs featured the enriched terms cell proliferation and differentiation, and ovule development. Collectively, these findings suggest that lncRNAs have distinct roles in the SC lineage and the non-SC lineage.

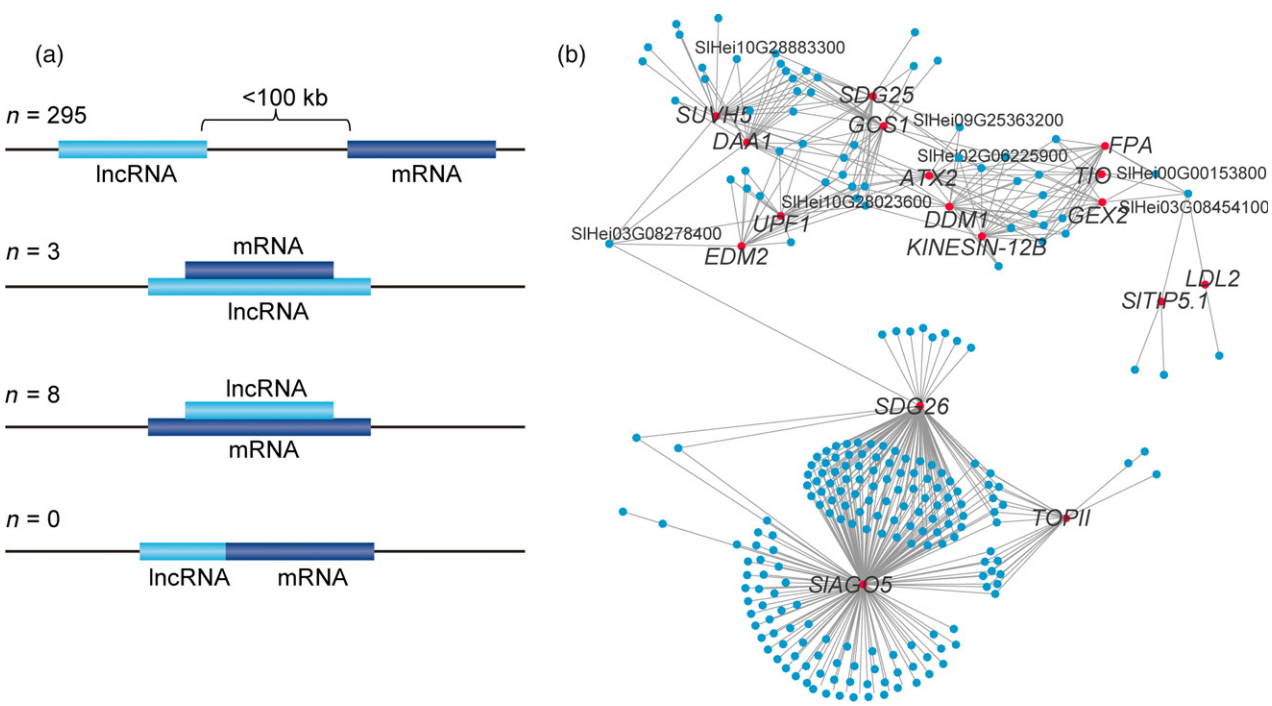


Figure 5. Co-expression network of mRNAs and long non-coding RNAs (lncRNAs).
(a) Graphical representation of co-expressed lncRNA and mRNA genes localized in a neighboring region (100 kb). The number is indicated for each case.
(b) A representative subnetwork visualized using Cytoscape with 17 generative cell (GC)-enriched and sperm cell (SC)-enriched mRNAs used as source nodes. Each red dot corresponds to a mRNA node and each blue dot to a key lncRNA node (>100 links, here showing only those links associated with this subnetwork). Representative lncRNA nodes are labeled with locus names. The lncRNAs in this subnetwork are highly correlated with epigenetic regulators (e.g. *SlAGO5*, *SDG26*, *SUVH5*, *DDM1*) and DUO1-targeted genes (e.g. *GCS1*, *DAA1*, *GEX2*).

Molecular markers distinguishing each cell type of the SC lineage

To explore the feasibility of using the transcriptomic data to identify molecular makers for each cell type of the SC lineage, we chose highly abundant (FPKM > 20) transcripts specifically expressed in each cell type of the SC lineage as candidates and visualized their expression by using single-molecule RNA-fluorescence *in situ* hybridization (FISH). Consistent with the RNA-seq data, RNA-FISH detected the lncRNA SlHei09G23622100 exclusively in the microspore; lncRNA SlHei07G20992500 and mRNA Solyc10g005480 (encoding a F-box family protein) were specifically in the GC; and lncRNA SlHei12G32629000 and mRNA Solyc06g076860 (encoding a cell division control protein) were found only in SCs. In contrast, the positive control ubiquitin (Solyc11g005670) was distributed in all three cell types of the SC lineage and tetrads (Figure 6, Table S8). These findings suggest that our data represent a resource for detecting further molecular markers distinguishing each cell type of this lineage.

DISCUSSION

The series of individual cells from the microspore to GC and finally SCs represents a well-characterized SC lineage. Dissecting the molecular program controlling SC lineage development is crucial for identifying the molecular markers recognizing each cell type of this lineage and understanding the molecular mechanisms for cell commitment, cell identity and fate establishment as well as SC specification in this process. We found sets of mRNAs and lncRNAs specifically expressed in each cell type. We further revealed molecular markers distinguishing these cell types by using single-molecule RNA-FISH in single cells. Our findings help resolve the distinctive identity of the microspore, the GC and SCs and provide a rich source of data for further delineating the mechanisms underlying SC lineage development.

The SC lineage has a unique landscape of lncRNAs

Although lines of evidence show that specific patterns of lncRNA expression coordinate cell state, differentiation

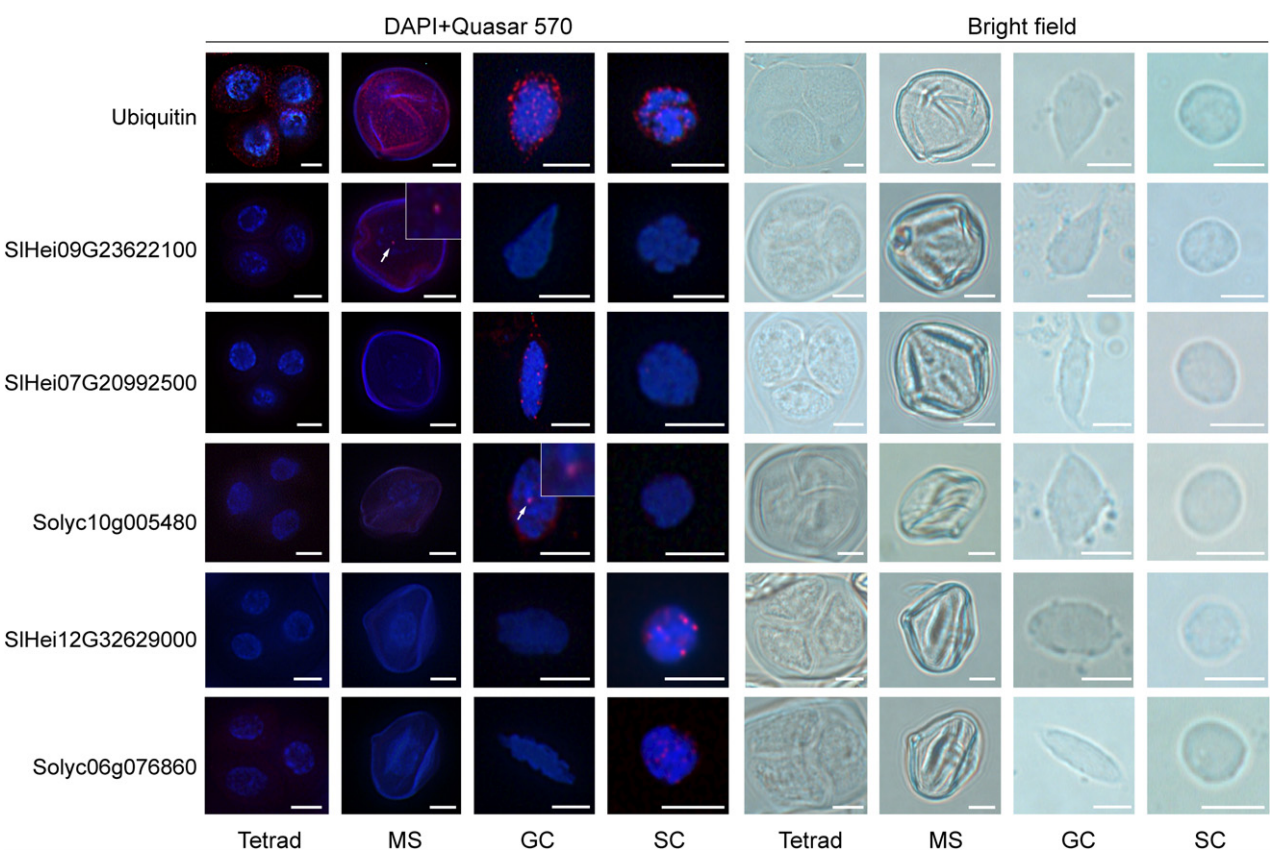


Figure 6. Single-cell and single-molecule fluorescence *in situ* hybridization (FISH) identification of molecular markers distinguishing the three cell types of the sperm cell (SC) lineage.
The indicated single-molecule signals in tetrad, microspore (MS), generative cell (GC) or SC were imaged by using 22-nucleotide DNA FISH probes with 3′ fusion to Quasar 570 dye. Cells were counterstained with 4′,6-diamidino-2-phenylindole. Scale bars = 5 μm. White arrows and magnified images indicate typical FISH spots.

and development (Quinn and Chang, 2016), transcriptional profiling of lncRNAs in the SC lineage is still lacking. We identified 53 651 lncRNA-coding genes in tomato genome by genome re-annotation, of which 90.47% were intergenic (lincRNAs). Similarly, 73% of lncRNAs (2224) identified in rice were lincRNAs (Zhang *et al.*, 2014). In contrast, among about 40 000 lncRNAs detected in Arabidopsis, most (>30 000) were long non-coding natural antisense transcripts, with only 6480 being lincRNAs (Liu *et al.*, 2012; Wang *et al.*, 2014), which implies that a larger genome size may have the potential to encompass more lincRNAs.

We found that 18 129 (56.8%) of 31 931 lncRNAs detected in the seven samples were enriched in the SC lineage. Likewise, a high proportion of lncRNAs were enriched in rice anthers and specifically expressed in human testis (Cabili *et al.*, 2011; Zhang *et al.*, 2014). Accordingly, our co-expression network revealed that lncRNAs enriched in the SC lineage and non-SC lineage showed co-expression with different sets of mRNAs, which were enriched for distinct GO terms characterizing major molecular and biological events of the corresponding cell/tissue type (for details see Results). This finding implies distinct functions of lncRNAs enriched in each cell type. Together, these results suggest that SC lineage-enriched lncRNAs might play essential roles in sex-cell differentiation and development. A study of mouse embryonic stem cells showed prevalent local correlation of gene expression between neighboring lncRNAs and mRNAs, which suggests extensive roles for lncRNAs in regulating gene expression in *cis* (Engreitz *et al.*, 2016). However, we found a small portion of lncRNAs to be co-expressed with their neighboring genes, a similar situation to that reported in rice (Zhang *et al.*, 2014), which implies that *trans*-acting lncRNAs may be more prevalent in plants.

Reprogramming gene expression and SC lineage development

The microspore shared more similarities in its gene expression profile with somatic cells than with its daughter the GC or SCs. The microspore also expressed regulators of shoot, root and cambium stem cells which play pivotal roles in maintenance of the somatic lineage. Consistently, other studies have shown extensive overlap between transcriptomes of the microspore and suspension cells in Arabidopsis and similarity between gene expression profiles of the microspore and callus in rice (Honys and Twell, 2004; Wei *et al.*, 2010). Retaining the somatic molecular features suggests that cell fate transition is not completed at the microspore stage. Together with the cytological observation that the microspore can be dedifferentiated and generates a somatic embryo *in vitro* (Segui-Simarro and Nuez, 2007; Daghma *et al.*, 2014; Seifert *et al.*, 2016), these findings indicate somatic fate and a potential pluripotent state of the microspore at the molecular level.

We showed that SC lineage development involves global repression of somatic and metabolic genes immediately following asymmetric division and coordinated upregulation of a small set of genes from the microspore to SCs or at the GC stage. Studies also showed significantly reduced complexity of the mRNome in SCs compared with the microspore in Arabidopsis and rice (Rutley and Twell, 2015). We found mRNAs associated with negative regulation of metabolism enriched in the GC; mRNAs conserved in SCs of tomato, rice and Arabidopsis were enriched for the GO terms negative regulation of transcription and metabolism, which implies that the repression of these processes is crucial for GC fate and SC specification. In mammals, primordial germ cells (PGCs) are highly specialized pre-meiotic precursors of sperms and eggs. The specification of PGCs relies on sequential suppression of somatic fate and activation of germ cell programming (Magnusdottir and Surani, 2014). Inhibition of somatic transcription ensures that PGCs escape from a somatic fate (Robert *et al.*, 2015). Therefore, we concluded that the global transcriptional repression switch from microspore to GC and to SCs, especially the global repression of somatic and metabolic genes, may define the identity of the GC and confer commitment of the GC to SCs. These genes upregulated from the microspore to SCs are implicated in gene silencing, cell wall organization and transporters, which are essential for SC specification and fertilization (Dong *et al.*, 2005; Lu *et al.*, 2011); those upregulated at the GC stage are implicated in the cell cycle and response to stress; examples are *DUO1* and its target genes which are required for GC division and SC specification (Borg *et al.*, 2011, 2014). At the genomic level, we also showed progressive repression of genes in repressive spots from the microspore to SCs and activation of genes in active spots, preferentially in the GC or from the microspore to the GC to SCs. Together, coordinated global repression of somatic and metabolic genes and activation of GC/SC-preferential genes are required for cell fate establishment and development of the SC lineage.

Synergistic regulation of reprogramming gene expression in the SC lineage

We revealed that reprogramming may be synergistically achieved by multiple regulators including TFs, hormone signaling and chromatin-related machineries. The transcriptionally repressed gene set included hormone signaling-related genes and more than 668 TFs, some of which have crucial roles in asymmetric cell division and identity determination (MacAlister *et al.*, 2007; Pillitteri *et al.*, 2007; Oh *et al.*, 2010), which suggests the involvement of these regulators in reprogramming. Furthermore, developmental specification of animal cells is associated with increasing chromatin restriction (Zhu *et al.*, 2013). Consistently, the microspore has dispersed chromatin, whereas the GC and

SCs have highly condensed chromatin. This observation agrees with the finding that the diversity of mRNAs and lncRNAs was significantly reduced from the microspore to the GC and finally to SCs.

We showed unique expression patterns of chromatin-related genes in the SC lineage: chromatin-related genes were generally expressed at higher levels in the microspore than in the GC and SCs, but a set of chromatin-related genes related to DNA and histone modifications, RNA-mediated gene silencing and chromatin organization had the highest expression in the GC and/or SCs among these examined cells and tissues. Considering the global transcriptional repression switch and increased chromatin condensation from the microspore to the GC to SCs (also see above), it is difficult to explain the unique expression patterns of chromatin-related genes. First, we cannot exclude that the microspore has inherited transcripts from pre-meiotic cells. Second, epigenetic signatures could have been established for several genes in the microspore to repress or activate genes immediately following asymmetric division. In mammals, the epigenetic marks–poised domains of somatic development genes in PGC precursors are transmitted to PGCs and subsequent mature gametes, and the poised state has been suggested to be important for repressing somatic gene expression and maintaining germ cell identity (Lesch and Page, 2014). Third, these GC- and SC-preferential epigenetic genes may represent a set of epigenetic factors that are important for reprogramming. These epigenetic pathways were enriched in mRNAs conserved in SCs across tomato, rice and Arabidopsis. We showed that expression of the active histone marks H3K4me3 and H3K9ac was high in the microspore but repressive H3K9 methylation was high in SCs. Therefore, the prolonged GC stage in tomato, compared with that of tricellular pollen, may not merely prepare for mitotic division but is a stage at which epigenetic modifications undergo a remarkable transition from domination of active marks to domination of repressive marks. Immunolabelling studies in mature *Quercus suber* pollen also support that the GC features a similar level of H3K9me2 and H3K4me3 modification (Ribeiro *et al.*, 2009). Because of the limited choice of specific antibodies we did not check other histone modification marks such as H3K4me2 and H3K27me3, whose levels seem to vary from species to species (Borg and Berger, 2015). Together with the finding that Pols IV and V, which are required for RdDM and gene silencing (Ream *et al.*, 2009; Haag and Pikaard, 2011), dispayed 'V-type' expression patterns from the microspore to the GC to SCs, these results suggest that histone marks, RdDM and RNA-mediated gene silencing seem to be important epigenetic mechanisms regulating transcriptional reprogramming.

The transcriptional apparatus in SCs has low activity

Previous inhibitor experiments suggested low activity of the transcriptional apparatus in the MPG (Mascarenhas, 1993; Honys and Twell, 2004), which is consistent with our finding of gradually decreased levels of Pols I, II and III in the developing SC lineage. The complexity of transcriptomes was significantly reduced from the microspore to the GC to SCs, but each cell type had preferential transcripts. These data suggest that the activity of the transcriptional apparatus is reduced in the GC and further in SCs. Furthermore, the GC and SCs had much lower levels of 25S rRNA and 40S ribosomal protein S14-1 than did the microspore. In human spermatozoa, 28S and 18S rRNA are conspicuously absent, possibly to ensure cessation of translation (Ostermeier *et al.*, 2002; Johnson *et al.*, 2011). Together, these lines of evidence indicate that SCs may have low activity of the transcriptional as well as the translational apparatus, unlike mammalian sperm, which shows cessation of transcription and translation.

EXPERIMENTAL PROCEDURES

Plant materials

The tomato cultivar Heinz 1706 (*Solanum lycopersicum*) was planted in a climate-controlled chamber under 14-h light/10-h dark at 28°C for 2 weeks, then fully expanded leaves were collected. One-month-old friable callus was generated from tomato hypocotyls on solid Murashige and Skoog medium containing 1 mg L^{-1} 2,4-dichlorophenoxyacetic acid and 0.5 mg L^{-1} kinetin in a 12-h light/12-h dark cycle at 25°C. Pistils were harvested 1 day before anthesis. The MPGs were collected from plants grown in a greenhouse. Generative cells were isolated from just-germinated pollen cultured *in vitro* for 20 min and SCs from 10-h-cultured pollen tubes (Lu *et al.*, 2015). For microspore isolation, 7-mm long flowers were excised and 5-mm long anthers were squashed in ice-cold 0.6 M mannitol solution. The released microspores was filtered through a 75-μm nylon mesh and collected by centrifugation at 200 ***g*** for 2 min.

Extraction of RNA and strand-specific RNA sequencing (ssRNA-seq)

Total RNA was extracted by use of the Ambion mirVana RNA isolation kit (Thermofisher, http://www.thermofisher.com/) according to the manufacturer's instructions and examined on 1% agarose gels. The RNA purity was checked using a NanoPhotometer spectrophotometer (Implen, https://www.implen.de/). The RNA concentration was measured using a Qubit RNA assay kit with the Qubit 2.0 fluorometer (Thermofisher). The integrity of the RNA was assessed using a RNA Nano 6000 assay kit with the Agilent Bioanalyzer 2100 system (Agilent Technologies, https://www.agilent.com/). For each sample replicate, rRNA was removed from total RNA preparations (2 μg) using an Epicentre Ribo-zero rRNA removal kit (Epicentre, http://www.epibio.com/). A similar amount of rRNA-depleted RNA across all 21 sample replicates was used for library construction. Strand-specific sequencing libraries were generated using the rRNA-depleted RNA and NEBNext Ultra directional RNA library prep kit for Illumina (NEB, https://www.neb.com/) following the manufacturer's recommendations. The libraries were qualified on

the Agilent Bioanalyzer 2100 system, sequenced on an Illumina Hiseq 2000 platform (https://www.illumina.com/) and 100-bp paired-end reads were generated.

Re-annotation of the tomato genome

The RNA-seq data from the 21 libraries and other public tomato RNA-seq datasets downloaded from NCBI were merged. In the merged RNA-seq dataset, redundant reads and low-quality reads were removed by Trinity kmer filtering. Then, the filtered dataset was assembled by using SOAPDenovo-Trans and treated as expressed sequence tag (EST) evidence. Protein-coding and non-coding RNA genes were predicted by an evidence-based approach (Liang *et al.*, 2009) combining all plant protein sequences from SwissProt, all eudicot cDNAs/ESTs and the assembled transcripts as evidence. After gene prediction, motifs and domains were annotated using InterProScan to search publicly available databases, including ProDom (http://prodom.prabi.fr/), PRINTS (www.bioinf.manchester.ac.uk/dbbrowser/PRINTS/), Pfam (http://pfam.xfam.org/), SMART (http://smart.embl-heidelberg.de/), PANTHER (http://www.pantherdb.org/), SUPERFAMILY (http://supfam.org/SUPERFAMILY/), PIR (http://pir.georgetown.edu/) and PROSITE (http://prosite.expasy.org/). The predicted genes with only EST evidence and encoding proteins <100 amino acids long were classified as potential lncRNA genes.

Identification of lncRNA genes

Because of the large amount of RNA-seq data used in the gene annotation analysis, we retrieved many non-coding genes. To classify those genes, three filters were used. First, tRNA, rRNA, small nucleolar RNA and microRNA genes were predicted using INFERNAL (Nawrocki and Eddy, 2013) with a search against the Rfam (v12) database; these genes were treated as short non-coding RNAs and removed. Second, genes with coding protein domains were discarded. Finally, the Coding Potential Calculator (CPC) was used to predict lncRNAs (Kong *et al.*, 2007). Annotation of lncRNA genes involved use of an in-house Perl script.

Analysis by RT-PCR and qPCR

Total RNA was reverse transcribed by using the SuperScript III first-strand synthesis system (Invitrogen, http://www.invitrogen.com/). A 1:20 dilution of cDNA was used for semi-quantitative PCR (RT-PCR) and real-time quantitative PCR (qPCR) analysis. Gene-specific primers for RT-PCR and qPCR were generated using NCBI Primer-BLAST (https://www.ncbi.nlm.nih.gov/tools/primer-blast/) and PRIMEREXRESS 2.0 software. Relative quantification ($2^{-\Delta\Delta CT}$ method) qPCR involved the use of SYBR green master mix (Thermofisher) according to the manufacturer's instructions. 18S ribosomal RNA was used as an internal control in RT-PCR and qPCR.

Immunoblot analysis

Mature pollen grains and microspores were homogenized in an ice-cold extraction buffer [100 mM 2-amino-2-(hydroxymethyl)-1,3-propanediol (TRIS)-HCl, 18% sucrose, 10 mM $MgCl_2$, 2% SDS, 1× protease inhibitor cocktail, 1 mM DTT, pH 8.0] with a benchtop homogenizer FastPrep-24 (MP Biomedicals, https://www.mpbio.com/). The GCs and SCs were lysed in extraction buffer by vigorous pipetting and sonication for 5 min. Cell homogenates/lysates were centrifuged for 15 min at 7000 ***g*** and 4°C, and the resulting supernatant was used for 4–20% SDS-PAGE and Western blotting. Protein concentration was determined with the BCA protein assay kit (Thermofisher). The primary antibodies used were anti-RNA Pol II (#05-623-AF488, Millipore, http://www.merckmillipore.com/), anti-S14-1(#AS122111, Agrisera, https://www.agrisera.com/), anti-H3K4me3(#AS163190, Agrisera), anti-H3K9ac (#9649, Cell Signaling Technology, https://www.cellsignal.com/), anti-H3K9me1(#07-450, Millipore), anti-H3K9me2(#AS163194, Agrisera), anti-H3K9me3(#13969, Cell Signaling Technology) and anti-H3(#AS10710, Agrisera).

Single-molecule RNA-FISH

For hybridization with the microspore, we collected flower buds of length 6–7 mm and isolated anthers. Anthers were cut into pieces with a razor blade and submerged into ice-cold phosphate buffered saline (PBS) buffer (137 mM NaCl, 2.7 mM KCl, 8 mM Na_2HPO_4, 2 mM KH_2PO_4, pH 7.4) supplemented with 3.7% formaldehyde. The microspores were released by gentle stirring and then filtered through a 60-μm nylon mesh. Cells were washed three times in PBS and collected by centrifugation at 500 ***g*** for 2 min. For hybridization with GCs and SCs, germinated pollen grains and pollen tubes were harvested as described (Lu *et al.*, 2015) and incubated in a 15.3% sucrose solution for 10 min at 28°C to release GCs and SCs, which were filtered through an 11-μm nylon mesh and collected by centrifugation at 850 ***g*** for 4 min. Cell pellets were suspended in ice cold fixation buffer [10 mM 2-(*N*-morpholine)-ethanesulfonic acid (MES)-KOH, 10 mM NaCl, 18% sucrose, 3.7% formaldehyde, pH 6.0] and incubated for 2 h at 4°C. After fixation, cells were washed once with PBS, dropped onto a poly-L-lysine slide, covered with a coverslip, and then submerged in liquid nitrogen for 2 min. The coverslip was removed and the slide was air dried and then permeabilized overnight in 70% ethanol. The RNA-FISH procedure was performed as described (Raj *et al.*, 2008). Briefly, cells were washed twice with wash buffer [10% deionized formamide, 2× sodium citrate buffer (SSC)] and covered with 50 μl of hybridization buffer (10% dextran sulfate, 2× SSC, 10% deionized formamide, 2 mM vanadyl ribonucleoside complex, 125 nM Stellaris probes labeled with Quasar570) and a coverslip for 24 h at 37°C in a dark humid chamber. The slides were submerged in wash buffer containing 0.1 μg ml^{-1} 4,6-diamidino-2-phenylindole for 1 h at room temperature. Finally, slides were washed twice with 2× SSC and covered with anti-fade mountant. Fluorescent images were captured using a Zeiss Axio Imager A1 (https://www.zeiss.com/) equipped with a 100× objective. Images were deconvolved with the Huygens essential program.

Expression abundance and tissue-specific score

Cuffdiff (v2.1.1) was used to calculate FPKM for both lncRNA and protein-coding genes in each sample (Trapnell *et al.*, 2010). The expression abundance and tissue-specific score of each gene were calculated after a pseudo-count of 1 was added to the original raw FPKM value for each gene, with $\log_2$ transformation and *z*-score normalization by the following formula:

$$V' = \frac{\log_2(V+1)}{\sum_{i=1}^{n}\log_2(v_i+1)} \quad (1)$$

where $V = (v_1, \ldots, v_n)$ is the original raw FPKM abundance estimation of the transcript and V' is the new normalized abundance estimation. Calculation of the tissue-specific score (Jensen–Shannon divergence) for each gene involved an entropy-based measure as described (Cabili *et al.*, 2011).

Gene Ontology analysis

To understand the biological functions of tomato mRNAs, a BLASTX search was performed against the *Arabidopsis thaliana*

protein database with a cutoff of >40% identity and E-value < 0.05. Tissue- and cell-enriched transcripts were obtained as described (Wei *et al.*, 2010), and the GO terms for these enriched transcripts were obtained by using the web tools agriGO and PlantGSEA (Du *et al.*, 2010; Yi *et al.*, 2013) with *A. thaliana* TAIR10 as a reference.

Principal components analysis and clustering

Principal components analysis was performed using the R prcomp function on FPKM of gene expression. The PCA graph was plotted using R scatterplot3d. For clustering of mRNAs and lncRNAs, a transcript (see Table S2) was retained if FPKM > 0 in at least one sample of SC lineage cells. Then the filtered data were $\log_2$(FPKM + 1) transformed and delivered to perform hierarchical clustering with correlation distance and Ward's method based on the R pheatmap package.

Gene co-expression network construction

The high-confidence RNA-seq-based gene co-expression network was constructed by integrating the predictions of three state-of-the-art inference algorithms: weighted correlation network analysis (WGCNA) (Langfelder and Horvath, 2008), a graphical Gaussian model (GeneNet) (Schäfer *et al.*, 2001) and bagging statistical network inference (BC3NET) (de Matos Simoes and Emmert-Streib, 2012). Network visualization involved the use of Cytoscape (Cline *et al.*, 2007).

Genomic regions activated/silenced in the SC lineage

To determine whether a specific genomic region was activated or silenced in SCs, we divided the genome (version SL2.5) into regions of 1 Mb then calculated the expressed gene diversity (lncRNAs and mRNAs) in each 1-Mb region. A genomic region was considered activated when the gene diversity in SCs was five-fold higher than in the GC or microspore. A genomic region was considered silenced when the gene diversity in SCs was five-fold lower than in the GC or microspore.

ACCESSION NUMBERS

Sequence data of the representative genes shown in this article can be found in the SGN/GenBank data libraries under the accession numbers listed in Table S9. Raw data files are available in the GEO database under accession number GSE117185.

ACKNOWLEDGMENTS

This work was supported by the Ministry of Science and Technology (grant nos 2013CB945101 and 2012CB910504).

CONFLICT OF INTEREST

The authors declare no conflicts of interest.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article.

Figure S1. Characteristics of long non-coding RNAs and mRNAs.

Figure S2. The microspore shares molecular features with somatic cells.

Figure S3. Global transcription restriction during the sperm cell lineage development.

Figure S4. Gene Ontology analysis of mRNA sets.

Figure S5. Comparison of mRNA expression in the mature pollen grain and sperm cell lineage.

Figure S6. Comparative analysis of sperm cell mRNomes (all expressed mRNAs).

Figure S7. Dynamics of long non-coding RNAs expressed in the developing sperm cell lineage.

Figure S8. Physical maps of silenced or activated genomic regions in the sperm cell lineage.

Figure S9. Expression of key long non-coding RNAs.

Table S1. Statistics of RNA-sequencing reads.

Table S2. Expression of mRNAs and long non-coding RNAs in indicated cells and tissues.

Table S3. Primers and results for quantitative PCR.

Table S4. Stage-enriched mRNAs and long non-coding RNAs.

Table S5. Pearson's correlation coefficient matrix of indicated cells and tissues.

Table S6. List of silenced or activated genomic regions in indicated cells and tissues.

Table S7. Confidence score for co-expression network of mRNAs and lncRNAs.

Table S8. List of single-molecule fluorescence *in situ* hybridization probes.

Table S9. Accession numbers.

REFERENCES

Anderson, S.N., Johnson, C.S., Jones, D.S., Conrad, L.J., Gou, X., Russell, S.D. and Sundaresan, V. (2013) Transcriptomes of isolated *Oryza sativa* gametes characterized by deep sequencing: evidence for distinct sex-dependent chromatin and epigenetic states before fertilization. *Plant J.* **76**, 729–741.

Berger, F. and Twell, D. (2011) Germline specification and function in plants. *Annu. Rev. Plant Biol.* **62**, 461–484.

Borg, M. and Berger, F. (2015) Chromatin remodelling during male gametophyte development. *Plant J.* **83**, 177–188.

Borg, M., Brownfield, L., Khatab, H., Sidorova, A., Lingaya, M. and Twell, D. (2011) The R2R3 MYB transcription factor DUO1 activates a male germline-specific regulon essential for sperm cell differentiation in Arabidopsis. *Plant Cell*, **23**, 534–549.

Borg, M., Rutley, N., Kagale, S. *et al.* (2014) An EAR-dependent regulatory module promotes male germ cell division and sperm fertility in Arabidopsis. *Plant Cell*, **26**, 2098–2113.

Borges, F., Gomes, G., Gardner, R., Moreno, N., McCormick, S., Feijo, J.A. and Becker, J.D. (2008) Comparative transcriptomics of Arabidopsis sperm cells. *Plant Physiol.* **148**, 1168–1181.

Brownfield, L., Hafidh, S., Borg, M., Sidorova, A., Mori, T. and Twell, D. (2009a) A plant germline-specific integrator of sperm specification and cell cycle progression. *PLoS Genet.* **5**, e1000430.

Brownfield, L., Hafidh, S., Durbarry, A., Khatab, H., Sidorova, A., Doerner, P. and Twell, D. (2009b) Arabidopsis DUO POLLEN3 is a key regulator of male germline development and embryogenesis. *Plant Cell*, **21**, 1940–1956.

Cabili, M.N., Trapnell, C., Goff, L., Koziol, M., Tazon-Vega, B., Regev, A. and Rinn, J.L. (2011) Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. *Genes Dev.* **25**, 1915–1927.

Cline, M.S., Smoot, M., Cerami, E. *et al.* (2007) Integration of biological networks and gene expression data using Cytoscape. *Nat. Protoc.* **2**, 2366–2382.

Daghma, D.E., Hensel, G., Rutten, T., Melzer, M. and Kumlehn, J. (2014) Cellular dynamics during early barley pollen embryogenesis revealed by time-lapse imaging. *Front. Plant Sci.* **5**, 675.

Ding, J., Lu, Q., Ouyang, Y., Mao, H., Zhang, P., Yao, J., Xu, C., Li, X., Xiao, J. and Zhang, Q. (2012) A long noncoding RNA regulates photoperiod-sensitive male sterility, an essential component of hybrid rice. *Proc. Natl Acad. Sci. USA*, **109**, 2654–2659.

Dong, X., Hong, Z., Sivaramakrishnan, M., Mahfouz, M. and Verma, D.P. (2005) Callose synthase (CalS5) is required for exine formation during microgametogenesis and for pollen viability in *Arabidopsis*. *Plant J.* **42**, 315–328.

Du, Z., Zhou, X., Ling, Y., Zhang, Z. and Su, Z. (2010) agriGO: a GO analysis toolkit for the agricultural community. *Nucleic Acids Res.* **38**, W64–W70.

Earley, K., Lawrence, R.J., Pontes, O., Reuther, R., Enciso, A.J., Silva, M., Neves, N., Gross, M., Viegas, W. and Pikaard, C.S. (2006) Erasure of histone acetylation by Arabidopsis *HDA6* mediates large-scale gene silencing in nucleolar dominance. *Genes Dev.* **20**, 1283–1293.

Engel, M.L., Chaboud, A., Dumas, C. and McCormick, S. (2003) Sperm cells of *Zea mays* have a complex complement of mRNAs. *Plant J.* **34**, 697–707.

Engreitz, J.M., Haines, J.E., Perez, E.M., Munson, G., Chen, J., Kane, M., McDonel, P.E., Guttman, M. and Lander, E.S. (2016) Local regulation of gene expression by lncRNA promoters, transcription and splicing. *Nature*, **539**, 452–455.

Gabory, A., Jammes, H. and Dandolo, L. (2010) The *H19* locus: role of an imprinted non-coding RNA in growth and development. *BioEssays*, **32**, 473–480.

Gou, X., Yuan, T., Wei, X. and Russell, S.D. (2009) Gene expression in the dimorphic sperm cells of *Plumbago zeylanica*: transcript profiling, diversity, and relationship to cell type. *Plant J.* **60**, 33–47.

Greb, T. and Lohmann, J.U. (2016) Plant stem cells. *Curr. Biol.* **26**, R816–R821.

Gupta, R.A., Shah, N., Wang, K.C. *et al.* (2010) Long non-coding RNA HOTAIR reprograms chromatin state to promote cancer metastasis. *Nature*, **464**, 1071–1076.

Haag, J.R. and Pikaard, C.S. (2011) Multisubunit RNA polymerases IV and V: purveyors of non-coding RNA for plant gene silencing. *Nat. Rev. Mol. Cell Biol.* **12**, 483–492.

Hennig, L. and Derkacheva, M. (2009) Diversity of polycomb group complexes in plants: same rules, different players? *Trends Genet.* **25**, 414–423.

Heo, J.B. and Sung, S. (2011) Vernalization-mediated epigenetic silencing by a long intronic noncoding RNA. *Science*, **331**, 76–79.

Honys, D. and Twell, D. (2004) Transcriptome analysis of haploid male gametophyte development in Arabidopsis. *Genome Biol.* **5**, R85.

Iwakawa, H., Shinmyo, A. and Sekine, M. (2006) Arabidopsis *CDKA;1*, a *cdc2* homologue, controls proliferation of generative cells in male gametogenesis. *Plant J.* **45**, 819–831.

Jeddeloh, J.A., Bender, J. and Richards, E.J. (1998) The DNA methylation locus *DDM1* is required for maintenance of gene silencing in Arabidopsis. *Genes Dev.* **12**, 1714–1725.

Johnson, G.D., Sendler, E., Lalancette, C., Hauser, R., Diamond, M.P. and Krawetz, S.A. (2011) Cleavage of rRNA ensures translational cessation in sperm at fertilization. *Mol. Hum. Reprod.* **17**, 721–726.

Kasahara, R.D., Maruyama, D., Hamamura, Y., Sakakibara, T., Twell, D. and Higashiyama, T. (2012) Fertilization recovery after defective sperm cell release in Arabidopsis. *Curr. Biol.* **22**, 1084–1089.

Kaya, H., Shibahara, K.I., Taoka, K.I., Iwabuchi, M., Stillman, B. and Araki, T. (2001) *FASCIATA* genes for chromatin assembly factor-1 in Arabidopsis maintain the cellular organization of apical meristems. *Cell*, **104**, 131–142.

Kim, H.J., Oh, S.A., Brownfield, L., Hong, S.H., Ryu, H., Hwang, I., Twell, D. and Nam, H.G. (2008) Control of plant germline proliferation by SCF(FBL17) degradation of cell cycle inhibitors. *Nature*, **455**, 1134–1137.

Kong, L., Zhang, Y., Ye, Z.Q., Liu, X.Q., Zhao, S.Q., Wei, L. and Gao, G. (2007) CPC: assess the protein-coding potential of transcripts using sequence features and support vector machine. *Nucleic Acids Res.* **35**, W345–W349.

Langfelder, P. and Horvath, S. (2008) WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics*, **9**, 559.

Lau, O.S. and Bergmann, D.C. (2012) Stomatal development: a plant's perspective on cell polarity, cell fate transitions and intercellular communication. *Development*, **139**, 3683–3692.

Lee, Y.R.J., Li, Y. and Liu, B. (2007) Two Arabidopsis phragmoplast-associated kinesins play a critical role in cytokinesis during male gametogenesis. *Plant Cell*, **19**, 2595–2605.

Lesch, B.J. and Page, D.C. (2014) Poised chromatin in the mammalian germ line. *Development*, **141**, 3619–3626.

Liang, C., Mao, L., Ware, D. and Stein, L. (2009) Evidence-based gene predictions in plant genomes. *Genome Res.* **19**, 1912–1923.

Liu, C., Lu, F., Cui, X. and Cao, X. (2010) Histone methylation in higher plants. *Annu. Rev. Plant Biol.* **61**, 395–420.

Liu, J., Jung, C., Xu, J., Wang, H., Deng, S., Bernad, L., Arenas-Huertero, C. and Chua, N.H. (2012) Genome-wide analysis uncovers regulation of long intergenic noncoding RNAs in Arabidopsis. *Plant Cell*, **24**, 4333–4345.

Lu, Y., Chanroj, S., Zulkifli, L., Johnson, M.A., Uozumi, N., Cheung, A. and Sze, H. (2011) Pollen tubes lacking a pair of K^+ transporters fail to target ovules in Arabidopsis. *Plant Cell*, **23**, 81–93.

Lu, Y., Wei, L. and Wang, T. (2015) Methods to isolate a large amount of generative cells, sperm cells and vegetative nuclei from tomato pollen for "omics" analysis. *Front. Plant Sci.* **6**, 391.

MacAlister, C.A., Ohashi-Ito, K. and Bergmann, D.C. (2007) Transcription factor control of asymmetric cell divisions that establish the stomatal lineage. *Nature*, **445**, 537–540.

Magnusdottir, E. and Surani, M.A. (2014) How to make a primordial germ cell. *Development*, **141**, 245–252.

Mascarenhas, J.P. (1993) Molecular mechanisms of pollen tube growth and differentiation. *Plant Cell*, **5**, 1303–1314.

de Matos Simoes, R. and Emmert-Streib, F. (2012) Bagging statistical network inference from large-scale gene expression data. *PLoS ONE*, **7**, e33624.

Mori, T., Kuroiwa, H., Higashiyama, T. and Kuroiwa, T. (2006) GENERATIVE CELL SPECIFIC 1 is essential for angiosperm fertilization. *Nat. Cell Biol.* **8**, 64–71.

Nawrocki, E.P. and Eddy, S.R. (2013) Infernal 1.1: 100-fold faster RNA homology searches. *Bioinformatics*, **29**, 2933–2935.

Nowack, M.K., Grini, P.E., Jakoby, M.J., Lafos, M., Koncz, C. and Schnittger, A. (2006) A positive signal from the fertilization of the egg cell sets off endosperm proliferation in angiosperm embryogenesis. *Nat. Genet.* **38**, 63–67.

Oh, S.A., Johnson, A., Smertenko, A., Rahman, D., Park, S.K., Hussey, P.J. and Twell, D. (2005) A divergent cellular role for the FUSED kinase family in the plant-specific cytokinetic phragmoplast. *Curr. Biol.* **15**, 2107–2111.

Oh, S.A., Park, K.S., Twell, D. and Park, S.K. (2010) The *SIDECAR POLLEN* gene encodes a microspore-specific LOB/AS2 domain protein required for the correct timing and orientation of asymmetric cell division. *Plant J.* **64**, 839–850.

Okada, T., Bhalla, P.L. and Singh, M.B. (2006) Expressed sequence tag analysis of *Lilium longiflorum* generative cells. *Plant Cell Physiol.* **47**, 698–705.

Ostermeier, G.C., Dix, D.J., Miller, D., Khatri, P. and Krawetz, S.A. (2002) Spermatozoal RNA profiles of normal fertile men. *Lancet*, **360**, 772–777.

Park, S.K., Howden, R. and Twell, D. (1998) The *Arabidopsis thaliana* gametophytic mutation *gemini pollen1* disrupts microspore polarity, division asymmetry and pollen cell fate. *Development*, **125**, 3789–3799.

Park, S.K., Rahman, D., Oh, S.A. and Twell, D. (2004) *gemini pollen 2*, a male and female gametophytic cytokinesis defective mutation. *Sex. Plant Reprod.* **17**, 63–70.

Paule, M.R. and White, R.J. (2000) Survey and summary: transcription by RNA polymerases I and III. *Nucleic Acids Res.* **28**, 1283–1298.

Penny, G.D., Kay, G.F., Sheardown, S.A., Rastan, S. and Brockdorff, N. (1996) Requirement for *Xist* in X chromosome inactivation. *Nature*, **379**, 131–137.

Pillitteri, L.J., Sloan, D.B., Bogenschutz, N.L. and Torii, K.U. (2007) Termination of asymmetric cell division and differentiation of stomata. *Nature*, **445**, 501–505.

Pumplin, N., Sarazin, A., Jullien, P.E., Bologna, N.G., Oberlin, S. and Voinnet, O. (2016) DNA methylation influences the expression of *DICER-LIKE4* isoforms, which encode proteins of alternative localization and function. *Plant Cell*, **28**, 2786–2804.

Quinn, J.J. and Chang, H.Y. (2016) Unique features of long non-coding RNA biogenesis and function. *Nat. Rev. Genet.* **17**, 47–62.

Raj, A., van den Bogaard, P., Rifkin, S.A., van Oudenaarden, A. and Tyagi, S. (2008) Imaging individual mRNA molecules using multiple singly labeled probes. *Nat. Methods*, **5**, 877–879.

Rajakumara, E., Law, J.A., Simanshu, D.K., Voigt, P., Johnson, L.M., Reinberg, D., Patel, D.J. and Jacobsen, S.E. (2011) A dual flip-out mechanism

for 5mC recognition by the Arabidopsis SUVH5 SRA domain and its impact on DNA methylation and H3K9 dimethylation in vivo. *Genes Dev.* **25**, 137–152.

Ream, T.S., Haag, J.R., Wierzbicki, A.T., Nicora, C.D., Norbeck, A.D., Zhu, J.-K., Hagen, G., Guilfoyle, T.J., Paša-Tolić, L. and Pikaard, C.S. (2009) Subunit compositions of the RNA-silencing enzymes Pol IV and Pol V reveal their origins as specialized forms of RNA polymerase II. *Mol. Cell*, **33**, 192–203.

Ribeiro, T., Viegas, W. and Morais-Cecilio, L. (2009) Epigenetic marks in the mature pollen of *Quercus suber* L. (Fagaceae). *Sex. Plant Reprod.* **22**, 1–7.

Rinn, J.L. and Chang, H.Y. (2012) Genome regulation by long noncoding RNAs. *Annu. Rev. Biochem.* **81**, 145–166.

Robert, V.J., Garvis, S. and Palladino, F. (2015) Repression of somatic cell fate in the germline. *Cell. Mol. Life Sci.* **72**, 3599–3620.

Russell, S.D., Gou, X., Wong, C.E., Wang, X., Yuan, T., Wei, X., Bhalla, P.L. and Singh, M.B. (2012) Genomic profiling of rice sperm cell transcripts reveals conserved and distinct elements in the flowering plant male germ lineage. *New Phytol.* **195**, 560–573.

Rutley, N. and Twell, D. (2015) A decade of pollen transcriptomics. *Plant Reprod.* **28**, 73–89.

Schäfer, J., Opgen-Rhein, R.S. and Strimmer, K. (2001) Reverse engineering genetic networks using the GeneNet package. *J. Am. Stat. Assoc.* **96**, 1151–1160.

Segui-Simarro, J.M. and Nuez, F. (2007) Embryogenesis induction, callogenesis, and plant regeneration by in vitro culture of tomato isolated microspores and whole anthers. *J. Exp. Bot.* **58**, 1119–1132.

Seifert, F., Bossow, S., Kumlehn, J., Gnad, H. and Scholten, S. (2016) Analysis of wheat microspore embryogenesis induction by transcriptome and small RNA sequencing using the highly responsive cultivar "Svilena". *BMC Plant Biol.* **16**, 97.

Skene, P.J. and Henikoff, S. (2013) Histone variants in pluripotency and disease. *Development*, **140**, 2513–2524.

Slotkin, R.K., Vaughn, M., Borges, F., Tanurdzic, M., Becker, J.D., Feijo, J.A. and Martienssen, R.A. (2009) Epigenetic reprogramming and small RNA silencing of transposable elements in pollen. *Cell*, **136**, 461–472.

Tomato Genome, C. (2012) The tomato genome sequence provides insights into fleshy fruit evolution. *Nature*, **485**, 635–641.

Trapnell, C., Williams, B.A., Pertea, G., Mortazavi, A., Kwan, G., van Baren, M.J., Salzberg, S.L., Wold, B.J. and Pachter, L. (2010) Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat. Biotechnol.* **28**, 511–515.

Wang, H., Chung, P.J., Liu, J., Jang, I.C., Kean, M.J., Xu, J. and Chua, N.H. (2014) Genome-wide identification of long noncoding natural antisense transcripts and their responses to light in Arabidopsis. *Genome Res.* **24**, 444–453.

Wei, L.Q., Xu, W.Y., Deng, Z.Y., Su, Z., Xue, Y. and Wang, T. (2010) Genome-scale analysis and comparison of gene expression profiles in developing and germinated pollen in *Oryza sativa*. *BMC Genom.* **11**, 338.

Woo, H.R., Pontes, O., Pikaard, C.S. and Richards, E.J. (2007) VIM1, a methylcytosine-binding protein required for centromeric heterochromatinization. *Genes Dev.* **21**, 267–277.

Xin, H.P., Peng, X.B., Ning, J., Yan, T.T., Ma, L.G. and Sun, M.X. (2011) Expressed sequence-tag analysis of tobacco sperm cells reveals a unique transcriptional profile and selective persistence of paternal transcripts after fertilization. *Sex. Plant Reprod.* **24**, 37–46.

Xu, H., Swoboda, I., Bhalla, P.L. and Singh, M.B. (1999) Male gametic cell-specific gene expression in flowering plants. *Proc. Natl Acad. Sci. USA*, **96**, 2554–2558.

Xu, C., Luo, F. and Hochholdinger, F. (2016) LOB domain proteins: beyond lateral organ boundaries. *Trends Plant Sci.* **21**, 159–167.

Yang, L., Liu, Z., Lu, F., Dong, A. and Huang, H. (2006) SERRATE is a novel nuclear regulator in primary microRNA processing in Arabidopsis. *Plant J.* **47**, 841–850.

Yi, X., Du, Z. and Su, Z. (2013) PlantGSEA: a gene set enrichment analysis toolkit for plant community. *Nucleic Acids Res.* **41**, W98–W103.

Yoine, M., Ohto, M.A., Onai, K., Mita, S. and Nakamura, K. (2006) The *Iba1* mutation of UPF1 RNA helicase involved in nonsense-mediated mRNA decay causes pleiotropic phenotypic changes and altered sugar signalling in Arabidopsis. *Plant J.* **47**, 49–62.

Zhang, Y.C., Liao, J.Y., Li, Z.Y., Yu, Y., Zhang, J.P., Li, Q.F., Qu, L.H., Shu, W.S. and Chen, Y.Q. (2014) Genome-wide screening and functional analysis identify a large number of long noncoding RNAs involved in the sexual reproduction of rice. *Genome Biol.* **15**, 512.

Zhu, J., Adli, M., Zou, J.Y. *et al.* (2013) Genome-wide chromatin state transitions associated with developmental and environmental cues. *Cell*, **152**, 642–654.

