Plant Reproduction (2019) 32:29–38
https://doi.org/10.1007/s00497-018-00355-4

METHODS PAPER

Transcriptomics of Arabidopsis sperm cells at single-cell resolution

Chandra Shekhar Misra[1,2] · Mário R. Santos[1] · Mariana Rafael-Fernandes[1] · Nuno P. Martins[1] · Marta Monteiro[1] · Jörg D. Becker[1]

Received: 23 October 2018 / Accepted: 12 December 2018 / Published online: 24 January 2019

Abstract

Key message **We present a detailed protocol for isolation of single sperm cells and transcriptome analysis to study variation in gene expression between sperm cells.**

Abstract Male gametophyte development in flowering plants begins with a microspore mother cell, which upon two consecutive cell divisions forms a mature pollen grain containing a vegetative nucleus and two sperm cells. Pollen development is a highly dynamic process, involving changes at both the transcriptome and epigenome levels of vegetative nuclei and the pair of sperm cells that have their own cytoplasm and nucleus. While the overall transcriptome of Arabidopsis pollen development is well documented, studies at single-cell level, in particular of sperm cells, are still lacking. Such studies would be essential to understand whether and how the two sperm cells are transcriptionally different, in particular once the pollen tube grows through the transmitting tissue of the pistil. Here we describe a detailed protocol for isolation of single sperm cells from growing pollen tubes and analysis of their transcriptome.

Keywords *Arabidopsis thaliana* · Sperm cell · Single-cell RNA-seq · Transcriptome · Semi in vivo pollen tube growth

Introduction

The male germline in flowering plants is formed when post-meiotic haploid microspores undergo an asymmetric division, producing a larger vegetative cell that exits the cell cycle, and a smaller generative cell (or male germ cell) that undergoes a second division to produce twin sperm cells (reviewed in Berger and Twell 2011). The result is a unique "cell within a cell system", combining two independent cell lineages with different fates (Twell 2011). While the vegetative cell controls pollen tube growth, the two sperm cells delivered by the pollen tube to the embryo sac are required for double fertilization.

Over the last decade, a wealth of transcriptome and methylome data from *Arabidopsis thaliana* pollen (Hony and Twell 2004; Becker et al. 2003; Pina et al. 2005) and sperm cells (Borges et al. 2008; Slotkin et al. 2009; Calarco et al. 2012; Ibarra et al. 2012) have improved our understanding of this unique system. These studies are complemented by transcriptome analyses of sperm cell samples from other angiosperms, e.g. from maize (Engel et al. 2003), rice (Anderson et al. 2013) and tomato (Liu et al. 2018). Importantly, studies in Arabidopsis have shown that pollen tubes undergo transcriptomic changes as they travel through the transmitting tissue of the pistil (Qin et al. 2009; Lin et al. 2014; Leydon et al. 2017), but to what extent the sperm cells contribute to these changes remains unknown. Though sperm cells are passively transported in the growing pollen tube (Zhang et al. 2017), changes in their transcriptome are to be expected, since sperm cells are believed to progress from S phase of the cell cycle at anthesis to G2 just before fertilization (Friedman 1999).

Sperm cells in most flowering plants are isomorphic. However, a few plant species have dimorphic sperm cells caused by

A contribution to the special issue 'Cellular Omics Methods in Plant Reproduction Research'.

Communicated by Thomas Dresselhaus.

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s00497-018-00355-4) contains supplementary material, which is available to authorized users.

✉ Jörg D. Becker
jbecker@igc.gulbenkian.pt

[1] Instituto Gulbenkian de Ciência, Rua da Quinta Grande 6, 2780-156 Oeiras, Portugal

[2] Instituto de Tecnologia Química e Biológica, Avenida da República, 2780-157 Oeiras, Portugal

an asymmetrical cell division of the generative cell (Mogensen 1992; Chen et al. 1995; Saito et al. 2002). Probably, the best-studied example is *Plumbago zeylanica*, where the larger sperm cell containing more mitochondria preferentially fuses with the central cell, and the smaller plastid-rich sperm cell fuses preferentially with the egg cell (Russell 1985). Differentially expressed genes between dimorphic sperm cells in Plumbago have been identified, but lacked clear homologs in Arabidopsis (Gou et al. 2009).

No study so far has addressed the question whether the isomorphic sperm cells in Arabidopsis differ at gene expression level. Such a study was not possible until now because of the technical challenges involved in collecting sperm cells to study their transcriptomes at single-cell level. But with the combination of sperm cell FACS (Borges et al. 2012; Santos et al. 2017) and single-cell RNA-seq (scRNA-seq) (Angermueller et al. 2016; Macaulay et al. 2016), it is now feasible to isolate single sperm cells and study the transcriptome profile of each sperm cell individually.

Single-cell analyses have been extensively performed using animal models to understand various aspects of biological processes, including gene expression dynamics and changes in the epigenetic landscape (Gawad et al. 2016; Macaulay et al. 2017). However, plant single-cell studies have been challenging due to the difficulties in obtaining single cells from plants. For the isolation of single cells from complex tissues, cell walls have to be removed by maceration or enzymatic digestion to obtain protoplasts (Efroni and Birnbaum 2016; Libault et al. 2017). Nevertheless, over the years several plant single-cell types have emerged as models, including root hairs (Libault et al. 2010; Qiao and Libault 2013), trichomes (Marks et al. 2008; Yang and Ye 2013), stomatal guard cells (Jin et al. 2013) and germline cells (pollen and egg cells) (reviewed in Rutley and Twell 2015). Plant single-cell-type isolation by shearing (e.g. trichomes, cotton fibres or root hairs) or other techniques without the need to prepare protoplasts (e.g. FACS and percoll gradients for pollen developmental stages) have accelerated the analysis of such plant cell types (Becker et al. 2014; Dupl'áková et al. 2016).

Here we show that using an optimized system for semi in vivo pollen tube growth coupled to FACS it is possible to collect individual sperm cells from growing pollen tubes. We confirmed the purity of the sorted sperm cells and were able to successfully analyse the transcriptome of single sperm cells by using an optimized scRNA-Seq protocol.

Materials and methods

Plant material and growth conditions

Arabidopsis thaliana wild type (Columbia-0) and a transgenic marker line harbouring *MGH3p::MGH3-eGFP* and *ACT11p::H2B-mRFP* (Borges et al. 2012) were used in this study. Plants were sown on soil, grown for 8 weeks in short-day conditions (8-h light at 21–23 °C) and then transferred to long-day conditions (16-h light) to induce flowering.

Semi in vivo pollen tube growth and sperm cell release

For semi in vivo pollen tube growth, the marker line was used to pollinate WT flower buds, emasculated 16 h prior to pollination. After 2 h, the pollinated pistil was excised and placed on double-sided tape. The excised pistil was then cut at the junction of style and ovary and placed gently on liquid Arabidopsis pollen germination medium, avoiding the submergence of the papillae (stigma) with pollen grains. The pistil was incubated for an additional 3–4 h for the pollen tubes to emerge from the cut end of the style (Boavida and McCormick 2007). The pollen tubes were burst using osmotic shock (8% w/v mannitol). The summary of the procedure is illustrated in Fig. 1.

After the pollen tubes were burst, the suspension was immediately transferred to ice-cold sperm extraction buffer: 1.3 mM H_3BO_3, 3.6 mM $CaCl_2 \cdot 2H_2O$, 0.74 mM KH_2PO_4, 438 mM Sucrose, 7 mM MOPS, 0.83 mM $MgSO_4 \cdot 7H_2O$, pH adjusted to 6.0 (Santos et al. 2017). In order to collect only intact cells by FACS, Sytox orange (25 nM final concentration) was added to the cell suspension and incubated for 10 min on ice (for details see Supplementary material 1).

Fluorescence activated cell sorting

Fluorescent activated cell sorting was carried out using a MoFlo Legacy (Beckman Coulter, Fort Collins, USA) with settings as described in Santos et al. 2017 (for details see Supplementary material 1). MoFlo settings for forward and side scatter were defined for GFP labelled sperm cells isolated from mature pollen grains. This gating strategy was used as a template for sperm cells obtained after bursting of semi in vivo grown pollen tubes. Single-cell suspension (sperm extraction buffer containing sperm cells, vegetative nuclei and debris) was then loaded onto the cell sorter and sorting was done using a 100-µm nozzle with sort mode set to single cell. Single cells were sorted into PCR strip tubes or 0.2 ml Eppendorf, prefilled with 2.5 µL of freshly prepared RLT plus lysis buffer supplemented with 1% β-mercaptoethanol. Doublets and dead/compromised cells were excluded based on forward scatter, pulse width and Sytox Orange fluorescence. Sorted single-cell lysates were vortexed, spun down and transferred into liquid nitrogen before being stored at − 80 °C. Tubes with 50 sperm cells or tubes with sorted single beads (Spherotech) were used as positive and negative controls, respectively.

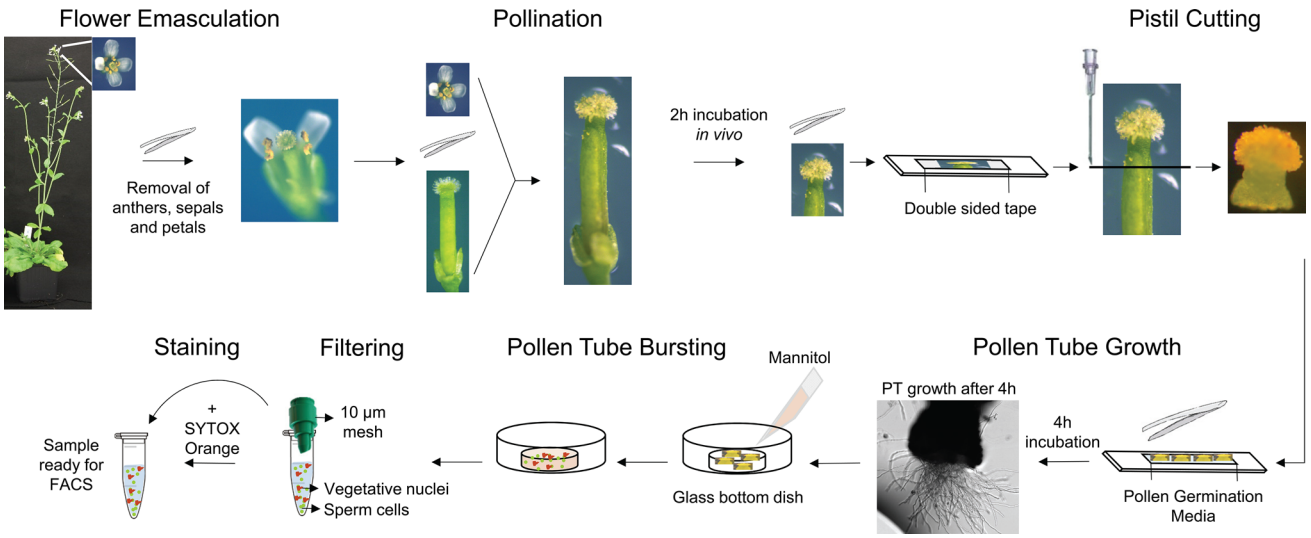

Fig. 1 Flow diagram of semi in vivo pollen tube growth assay and isolation of sperm cells

Single-cell RNA sequencing: reverse transcription

A detailed list of reagents and a step-by-step protocol are provided in Supplementary material 1. In brief, single-sorted sperm cells were subjected to Smart-seq2 as described earlier (Macaulay et al. 2016), with some modifications. A modified bead capture protocol was used, in which Dyna Beads MyOne Streptavidin C1 (Invitrogen) were washed and attached to oligo (dT) primers (5′AAGCAGTGGTATCAACGCAGAGTACTTTTTTTTTTTTTTTTTTTTTTTTTVN-3′, Biomer). Ten microlitres of conjugated beads was then mixed with thawed single-cell lysates and incubated at room temperature for 20 min with constant mixing at 2000 rpm. The tubes were then moved onto a ring magnet (Alpaqua, USA) for 1 min. Supernatant containing gDNA was discarded. The beads were washed twice with gDNA buffer, with care being taken to avoid loss of beads during washes.

Beads in each tube were then resuspended with 5 µl of RT mix containing: 0.5 µl of 10 mM dNTP (Thermo Fisher Scientific), 1.14 µl of nuclease-free water, 0.06 µl RNAse Inhibitor (40 U/µl, Takara), 1 µl of 5× Superscript II first-strand buffer (Thermo Fisher Scientific), 0.25 µl of 100 mM DTT (Invitrogen),1 µl of 5 M Betaine (Sigma-Aldrich), 0.3 µl of 100 mM $MgCl_2$ (Sigma-Aldrich), 0.5 µl of 10 µM TSO (5′-AAGCAGTGGTATCAACGCAGAGTACrGrG + G-3′, Exiqon), and 0.25 µl of Superscript II reverse transcriptase (200 U/µl, Thermo Fisher Scientific).

Each tube with RT mix was mixed well with the beads for 1 min at 2000 rpm. Reverse transcription was carried out by incubating the tubes at 42 °C for 60 min (template switching) followed by 50 °C for 30 min (template switching and RT) and finally enzyme inactivation at 60 °C for 10 min.

Single-cell RNA sequencing: PCR pre-amplification and cDNA purification

7.5 µl of PCR mix containing 6.25 µL of 2× KAPA HiFi Hot Start 6 Ready Mix (KAPA Biosystems), 0.125 µL of 10 µM ISPCR primer (5′-AAGCAGTGGTATCAACGCAGAGT-3′, IDT) and 1.125 µL nuclease-free water were added to each PCR tube for a final PCR volume of 12.5 µL.

The reaction was carried out with an initial incubation at 98 °C for 3 min followed by amplification using 21 cycles (98 °C for 20 s, 67 °C for 15 s, 72 °C for 6 min) and final extension at 72 °C for 5 min. PCR products were then cleaned by mixing them with an equal volume of Agencourt AMPure XP SPRI beads (Beckman Coulter) followed by incubation at room temperature for 8 min. The tubes were placed onto a magnet for 5 min before removing the supernatant. Beads were washed twice with 100 µl of 80% freshly prepared ethanol, with care being taken to avoid loss of beads during the washes. All traces of residual ethanol were removed, and beads were left to dry at room temperature for 5 min. The beads were then resuspended in 17.5 µl of molecular biology-grade water and incubated at room temperature for 5 min. The tubes were placed on the magnet for 2 min prior to transferring the supernatant containing the amplified cDNA to new Eppendorf tubes. The concentration of amplified cDNA was measured using a Qubit dsDNA high-sensitivity Assay kit (Thermo Fisher Scientific). Some single-cell cDNA libraries were then randomly selected to confirm size distribution using a NGS high-sensitivity kit on a Fragment Analyser (AATI).

RT-PCR analysis of sperm-specific transcripts

Quality and purity of single-cell RNA-seq libraries prepared by Smart-seq2 were verified by PCR using primers for the sperm-specific genes: *MGH3* (AT1G19890) and *DAZ3* (AT4G35700), reported to be highly expressed in sperm cells (Borges et al. 2008). *VEX1* (AT5G62850)-specific primers were used to control for contamination with vegetative nuclei (Borges et al. 2012). For this, the beads used in the previous step were incubated for a second time with 5 µl of molecular biology-grade water and the eluate used as a template in PCR of 35 cycles. See table S1 in Supplementary material 3 for list of primers.

Single-cell RNA sequencing: library preparation

Based on the results of RT-PCR, only those samples which showed amplification for sperm-specific genes were further processed for library preparation and sequencing. Libraries were prepared using a low-volume Nextera transposase-based protocol described in Baym et al. (2015). 2 nM of each library was pooled and sequenced on an in-house Illumina Nextseq 500 instrument, using 75-bp single-end reads.

RNA-seq data processing

RNA sequencing data from single sperm cells and 50 sperm cells were processed to check for quality of reads and alignment. For each RNA-seq data set, raw RNA-seq reads were checked for quality metrics with fastqc, and polyAs and adaptors were trimmed with cutadapt (Martin 2011) and trimmomatic (Bolger et al. 2014). The reads were then aligned to the Arabidopsis reference genome with recent annotation using Hisat 2 (Kim et al. 2015). HT-seq count (Ander et al. 2015) was then used to count gene features using recent annotation from Araport 11 (Cheng et al. 2017), and the expression matrix of raw counts was generated to be used for further analysis (Ander et al. 2015). Transcript quantification was achieved using Salmon (Patro et al. 2017). A transcript was considered to be expressed if transcript per million (TPM) was more than 1.

Results

Semi in vivo pollen tube growth in liquid medium

For FACS-based collection of single sperm cells, pollen tubes were first grown through pistils. Earlier studies using such semi in vivo assays (Qin et al. 2009) were mostly based on agar solidified medium. Here, however, the use of liquid pollen germination medium was important to allow the efficient collection of sperm cells for FACS. This optimized system for semi in vivo pollen tube (SIVPT) growth differs from what was originally described (Qin et al. 2009; Palanivelu and Preuss 2006): First, we pollinated the pistil in vivo and let the pollen tubes enter the style for about 2 h (Fig. 1). After that, the pistil was excised and placed on pollen germination medium for 4 h for the pollen tubes to grow and emerge from the cut end of the style. These modifications were incorporated to capture the transcriptome of sperm cells from SIVPT under the most natural conditions possible.

The emerging pollen tubes grew efficiently into the liquid pollen germination medium after 4 h, as shown in Fig. 2a. Moreover, our method of SIVPT growth on liquid germination medium allowed for the bursting of pollen tubes by osmotic shock and efficient release of intact sperm cells and vegetative nuclei as shown in Fig. 2b. The solution with released sperm cells and vegetative nuclei (occasionally still as male germ unit as seen in Fig. 2b) was collected and subjected to FACS.

Flow cytometry and sorting of single sperm cells

For FACS sorting, Arabidopsis sperm cells expressing GFP were isolated from mature pollen grains and used to establish the gating strategy (Fig. 2c), which then served as a template for FACS of sperm cells from pollen tubes (Fig. 2d). This pre-gating strategy using sperm cells from mature pollen grains was important, because the number of sperm cells collected from pollen tubes would not be sufficient to set the gate for sorting.

The solution containing the released sperm cells from growing pollen tubes was transferred to sperm extraction buffer and used for FACS. Sytox orange was used to distinguish intact sperm cells from dead/compromised cells. Only intact GFP positive cells, identified by GFP signal and absence of sytox-derived fluorescence, were selected during sorting. Forward and side scatter profiles were used to further discriminate sperm cells from debris. Based on this strategy, only a homogeneous population of sperm cells were sorted. In conclusion, we were able to successfully sort this rare population of sperm cells obtained from growing pollen tubes and use it for further transcriptome analysis.

Single sperm cell transcriptome

We developed an optimized protocol for single sperm cell RNA-seq by modifying the Smart-seq2 protocol (Picelli et al. 2014; Macaulay et al. 2016). Optimization of Smart-seq2 was done with single sperm cells along with bulk samples (e.g. 50 sperm cells) in order to assess to which extent a single-cell transcriptome would be different from bulk samples with respect to number of transcripts detected, alignment rate, etc.

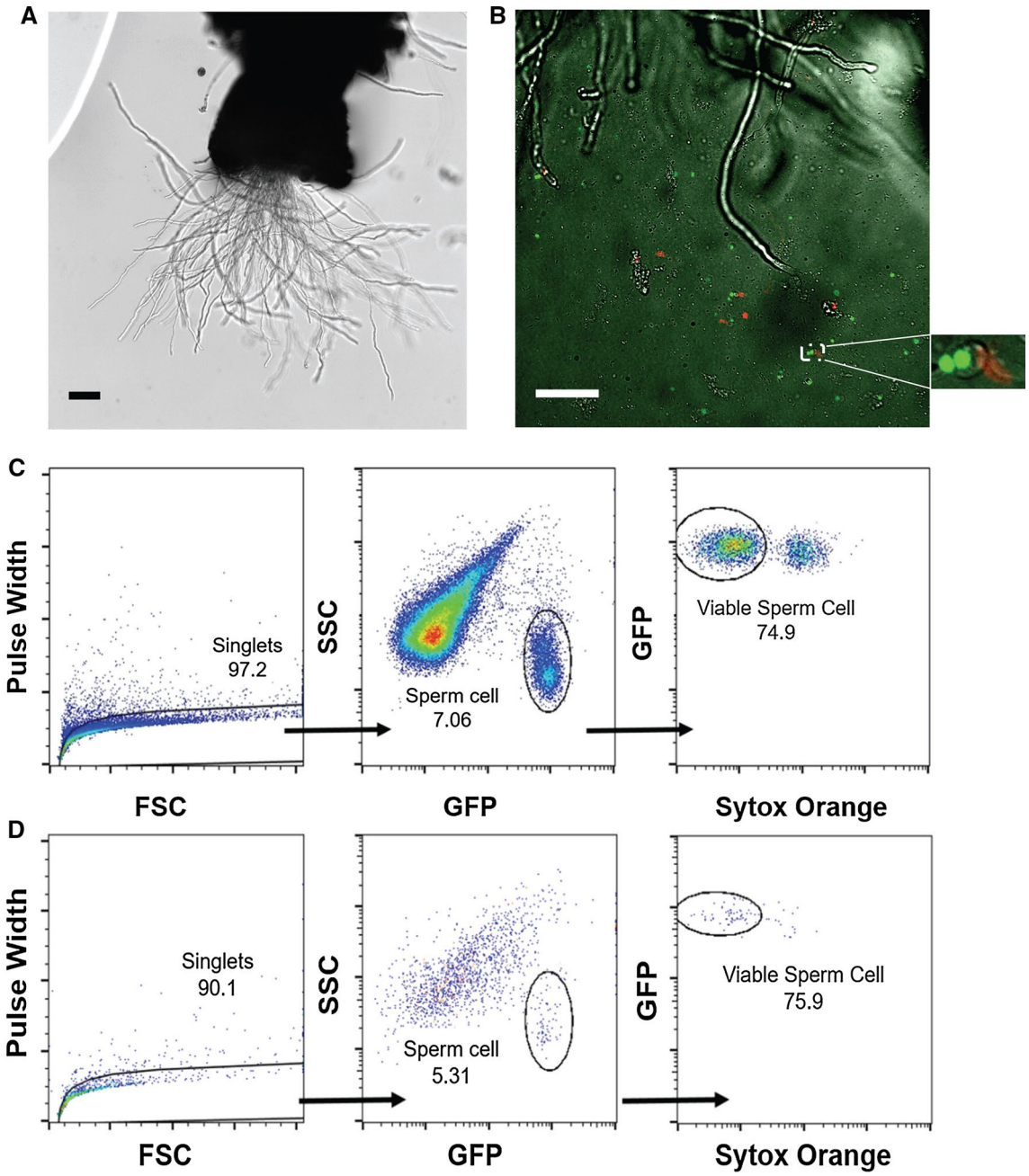

Fig. 2 Isolation of sperm cells from semi in vivo grown pollen tubes. **a** A cut pistil with pollen tubes emerging. Robust pollen tube growth was seen after 4 h. **b** GFP (green) labelled sperm cells and RFP (red) labelled vegetative nuclei floating in liquid medium after bursting of pollen tubes using 8% mannitol solution. **c**, **d** Flow cytometry profile of sperm cells for FACS: **c** Gating strategy template established using sperm cells from mature pollen grains, **d** Gating strategy used for sorting sperm cells from pollen tubes. Populations of gated sperm cells are marked by circles. *SSC* side scatter, *FSC* forward scatter. Bar: *A*-100 µm, *B*-50 µm

Single-cell cDNA was quantified and was found to have a concentration of around 0.3–0.7 ng/µl, with size distribution peaking around 1.5–1.7 kb (Fig. 3a–b). Since these two parameters were not a sufficient indicator of sperm specificity of the library, we tested if amplification of sperm-specific genes (*DAZ3, MGH3*) could be a viable alternative. To limit the loss of cDNA for subsequent steps, an additional step of re-elution of cDNA after the first beads clean-up was introduced and this extra eluted cDNA was used as template for PCR.

As shown in Fig. 3e, we were successful in performing PCR on the cDNA obtained from single cells and thus could add this important quality control step before library preparation to our protocol. To check for

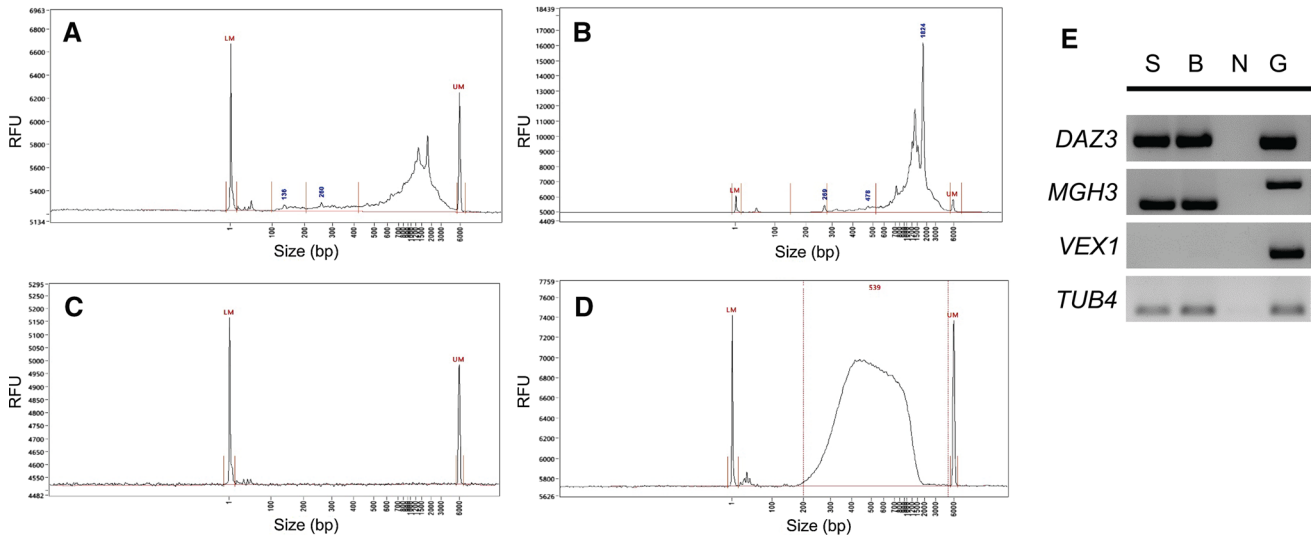

Fig. 3 Representative electropherogram plots from Fragment Analyzer showing various stages of single-cell RNA-seq library preparation. **a** Example of a good cDNA amplification from single cell with fragment distribution between 600 and 2500 bp, and an average size of 1500 to 2000 bp. **b** Example of a good cDNA amplification from bulk sample used as a positive control, with a similar fragment distribution. **c** Example of a negative (no cell) control. **d** Example of a typical tagmented and indexed Nextera library with a size distribution of 300–800 bp. RFU, relative fluorescence units. **e** RT-PCR analysis of sperm-specific transcripts (*DAZ3*, *MGH3*) from independent cDNA libraries obtained from single sperm cells and bulk populations. *VEX1* gene-specific primers were used as a reference to show purity (no vegetative nuclei contamination). *S* single sperm cell, *B* bulk population (50 sperm cells), *N* negative control, *G* genomic DNA. *TUB4* was used a positive control

contamination from vegetative nuclei, we analysed some of our cDNA libraries using gene-specific primers for *VEX1* (details in Supplementary material 1). Only single-cell cDNAs that showed PCR amplification for *DAZ3* were then used for Nextera library preparation (Fig. 3d).

Number, quality and mapping rate of single cells

We sequenced 80 single-cell libraries along with two bulk samples. Our single sperm cell RNA-seq protocol produced high-quality libraries with uniform coverage along the transcript length, consistent with the data obtained from bulk population (Fig. 4, S1 in Supplementary material 2).

The sperm cell libraries sequenced showed an average of 81% mapping rate with maximum reaching 88% (unpublished results), with large number of reads falling into exonic regions (Fig. S2 in Supplementary material 2). We also estimated the number of transcripts detected for the single-cell libraries using the Salmon tool, and the representative result for one single-cell and bulk (50 SC) library is shown in Fig. 5a. A transcript was considered to be expressed, if the transcript per million (TPM) was higher than 1. Our results from single sperm cells suggest that on average around 1900 ± 571 active genes were detected in each sperm cell.

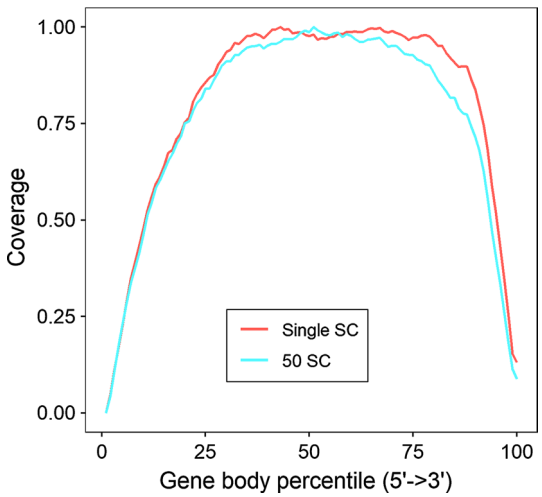

Fig. 4 Gene body coverage. 3′ and 5′ coverage of sequenced reads from 50 sperm cells and a single sperm cell along the length of transcripts. Graph was generated using geneBody_coverage.py from the RSeQC package version 2.6.4 (Wang et al. 2012) and reshaped using ggplot2 package from Rstudio (http://www.rstudio.com/)

Comparison of single-cell vs bulk samples

In order to assess whether the single-cell RNA-seq (scRNA-seq) data were in agreement with the results from bulk experiments, the scRNA-seq protocol was performed with a sample containing a bulk population of 50 sperm

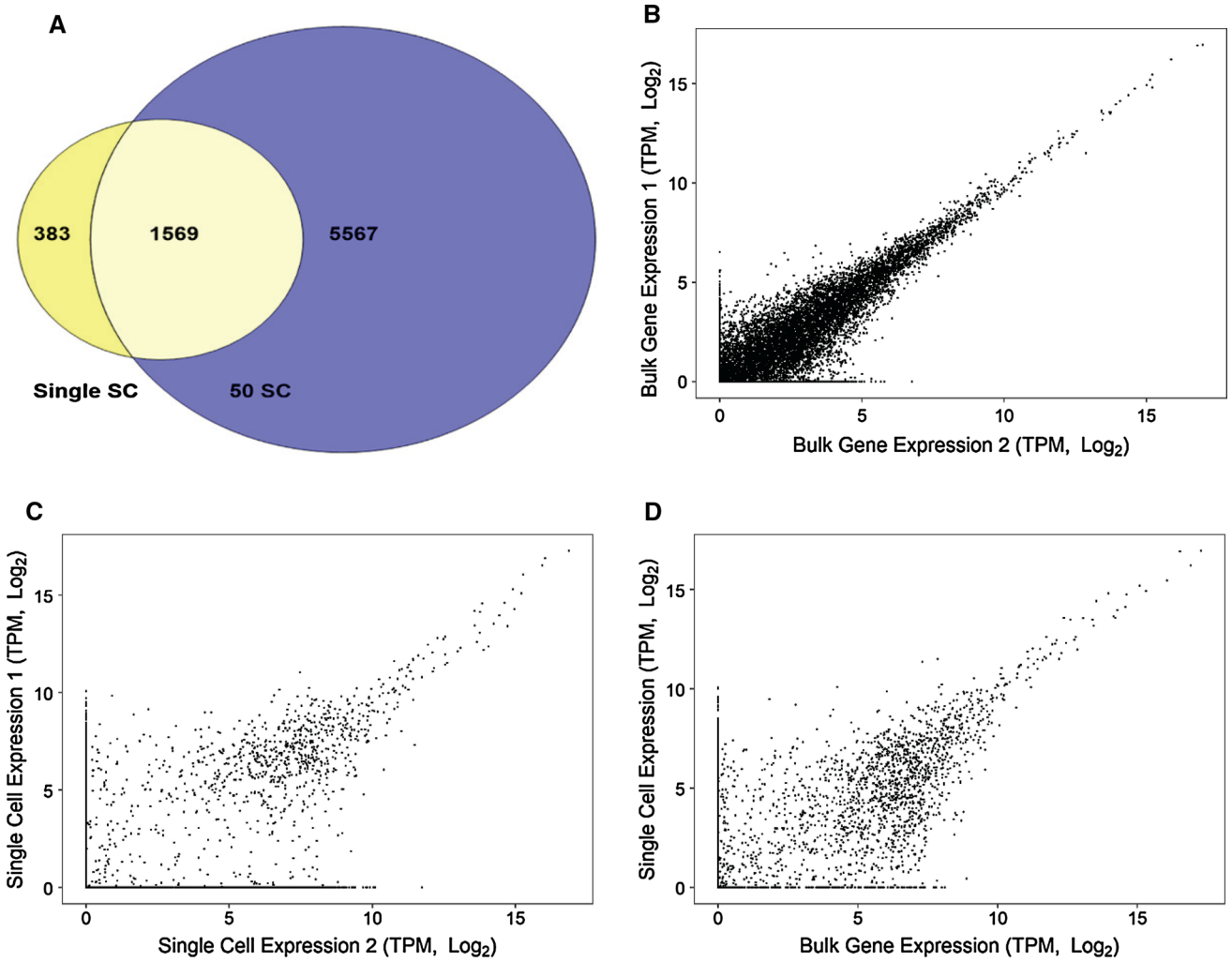

Fig. 5 Single-cell vs bulk gene expression. **a** Comparison of gene expression levels of representative single sperm cell (SC) and bulk sample (50 sperm cells). Number of transcripts detected above threshold (TPM > 1) is given. **b** Gene expression correlation (normalized) between two bulk samples (Pearson's coefficient $r = 0.84$); **c** correlation between two single cells (Pearson's coefficient $r = 0.57$); **d** correlation between bulk sample and gene expression of a single cell (Pearson's coefficient $r = 0.61$)

cells (SCs), which was processed in the same batch as single cells and sequenced together. As shown in Fig. 5a, the number of transcripts detected (Transcript per million, TPM > 1) was higher in bulk samples (50 SC) than in single sperm cells.

Gene expression data of bulk and single cells were used to draw scatter plots using R studio. Analysis of the data showed very high correlation between RNA-seq data obtained from two bulk populations, with Pearson correlation coefficient of 0.84 (Fig. 5b). This correlation dropped to 0.61 when comparing single-cell vs bulk population (Fig. 5c) and 0.57 when comparing two single-cell transcriptomes (Fig. 5d).

Discussion

Plant male germline development has been an active area of research, with a wealth of transcriptomic and epigenomic data available (see review Rutley and Twell 2015). All transcriptome studies on pollen (e.g. Hony and Twell 2004; Pina et al. 2005) and sperm development (Borges et al. 2008) so far have been performed using bulk cell isolations, thus obscuring any potential differences at single-cell level. Recent improvements of scRNA-seq protocols, a decrease in cost of high-throughput sequencing and advances in FACS isolation of Arabidopsis sperm

cells, however, have allowed us to develop a single sperm cell RNA-seq method that should enable us to address the question, if isomorphic Arabidopsis sperm cells differ in their transcriptomes. Rather than isolating sperm cells from mature pollen grains, which can be achieved in high numbers using existing protocols (Santos et al. 2017; Chumak et al. 2015), we optimized their isolation from semi in vivo grown pollen tubes. It is well known that pollen–pistil interactions elicit changes in the pollen tube transcriptome (Qin et al. 2009; Leydon et al. 2017), but to which extent changes in the sperm cell transcriptome contribute to this observation could not be assessed until now. Our optimized protocol for collection of sperm cells from semi in vivo growing pollen tubes removes these limitations.

Single-cell RNA-seq in general and Smart-seq2 in particular is a method primarily developed for mammalian cells that are much larger (10–100 µm), and thus assumingly with a higher cellular content (including RNA) than Arabidopsis sperm cells with a size of ~ 2.5 µm and very little cytoplasm. Therefore, there was no guarantee that we would obtain suitable cDNA libraries from single sperm cells of Arabidopsis. A number of intermediate optimization and quality control steps addressed this question. At the cDNA level we introduced a QC step, in which we performed a PCR-based validation of sperm cell specificity and purity (Fig. 3e). We checked selected single-cell libraries for possible contamination from the vegetative cell using primers specific for the VEX1 gene. However, VEX1 is not expressed at very high levels in pollen and it might be useful to check other highly expressed vegetative cell-specific genes. The PCR-based validation step together with cDNA fragmentation profiles allowed us to select only those cDNA libraries for further processing and sequencing that were of highest quality. Furthermore, we noticed that addition of 1% β-mercaptoethanol in the lysis buffer increased the alignment rate of RNA-seq data from single sperm cells by up to 40%. A growing number of recent single-cell studies support this observation (Villani et al. 2017; Puram et al. 2017).

We compared the transcriptomes of single sperm cells with those of 50 sperm cells (bulk sample). In the bulk sample, we detected more than 7000 genes (TPM > 1), which is 20% more genes than the 5829 genes detected using ATH1 gene arrays (Borges et al. 2008). Such an increase was expected though, given that the array did not cover all genes of the Arabidopsis genome. The results show, however, that our protocol is well suited to produce high-quality transcriptome data from very low number of cells. The number of genes detected in a single cell, however, is lower, with an average of around 1900 genes (Fig. 5a). These differences between single-cell and bulk libraries are mostly due to genes with low expression levels falling below the detection limit in single-cell expression data (Marinov et al. 2014). Technical variation in single-cell data may arise due to low genome coverage, lower sensitivity of single-cell protocols and high amplification bias, which also might explain how some genes are only expressed in single-cell but not in bulk data (Fig. 5a). This may lead to masking of real biological variation that may exist between single cells (Mantsoki et al. 2016). Our data also reveal that though strong correlation exists (Fig. 5b) between gene expression data of two bulk populations ($r = 0.84$), this correlation drops when we compare two single cells or single-cell expression with bulk population (Fig. 5c, d). How much of the variation observed in our single-cell data can be attributed to technical variation and how much represents true biological variation, possibly between sperm cells in the same male germ unit, remains to be determined.

The protocol presented here for isolation of single sperm cells from semi in vivo grown pollen tubes can be adapted to single-cell high-throughput platforms, allowing the analysis of higher cell numbers for a lower cost per cell, as long as these platforms support the capture of relatively small cells. Moreover, sperm cell extraction and stability under suitable buffer conditions need to be evaluated as most of these high-throughput scRNA-seq platforms are optimized for PBS buffer. With the number of applications for single-cell analysis increasing, this will open new avenues to study the transcriptome and epigenome of plant sperm cells on a single-cell level.

Author contribution statement JDB conceived the project. CSM, MRS and JDB designed the experiments for collection of single sperm cells using FACS. CSM optimized the protocol for pollen tube bursting and Smart-seq2 RNA-seq library preparation. MM, MF and NPM performed the FACS isolation using MoFlo and helped with quality control after FACS. CSM analysed the scRNA-seq data. CSM and JDB discussed the results and wrote the manuscript. All authors read and approved the manuscript.

Acknowledgements The authors would like to acknowledge the Genomics Unit and Bioinformatics Unit of IGC. The Genomics Unit was partially supported by ONEIDA Project (LISBOA-01-0145-FEDER-016417) co-funded by FEEI—"Fundos Europeus Estruturais e de Investimento" from "Programa Operacional Regional Lisboa 2020" and by national funds from FCT—"Fundação para a Ciência e a Tecnologia". This work was funded by FCT ERA-CAPS project EVOREPRO (ERA-CAPS-0001-2014). CSM acknowledges a doctoral fellowship from FCT (PD/BD/114362/2016) under the Plants for Life PhD Program. JDB received salary support from FCT through an "Investigador FCT" position.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

References

Anders S, Pyl PT, Huber W (2015) HTSeq—a Python framework to work with high-throughput sequencing data. Bioinformatics 31:166–169

Anderson SN, Johnson CS, Jones DS, Conrad LJ, Gou X, Russell SD, 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

Angermueller C, Clark SJ, Lee HJ, Macaulay IC, Teng MJ, Hu TX, Krueger F, Smallwood SA, Ponting CP, Voet T, Kelsey G (2016) Parallel single-cell sequencing links transcriptional and epigenetic heterogeneity. Nat Methods 13:229–232

Baym M, Kryazhimskiy S, Lieberman TD, Chung H, Desai MM, Kishony R (2015) Inexpensive multiplexed library preparation for megabase-sized genomes. PLoS One 10:e0128036

Becker JD, Boavida LC, Carneiro J, Haury M, Feijó JA (2003) Transcriptional profiling of Arabidopsis tissues reveals the unique characteristics of the pollen transcriptome. Plant Physiol 133:713–725

Becker JD, Takeda S, Borges F, Dolan L, Feijó JA (2014) Transcriptional profiling of Arabidopsis root hairs and pollen defines an apical cell growth signature. BMC Plant Biol 14:197

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

Boavida LC, McCormick S (2007) Technical Advance: temperature as a determinant factor for increased and reproducible in vitro pollen germination in *Arabidopsis thaliana*. Plant J 52:570–582

Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30:2114–2120

Borges F, Gomes G, Gardner R, Moreno N, McCormick S, Feijó JA, Becker JD (2008) Comparative transcriptomics of Arabidopsis sperm cells. Plant Physiol 148:1168–1181

Borges F, Gardner R, Lopes T, Calarco JP, Boavida LC, Slotkin RK, Martienssen RA, Becker JD (2012) FACS-based purification of Arabidopsis microspores, sperm cells and vegetative nuclei. Plant Methods 8:44

Calarco JP, Borges F, Donoghue MTA, Van Ex F, Jullien PE, Lopes T, Gardner R, Berger F, Feijó JA, Becker JD, Martienssen RA (2012) Reprogramming of DNA methylation in pollen guides epigenetic inheritance via small RNA. Cell 151:194–205

Chen GH, Hu ZM, Guo FL, Guan XL (1995) Male gametophyte development in Plumbago zeylanica: cytoplasm localization and cell determination in 708 the early generative cell. Protoplasma 186:79–86

Cheng CY, Krishnakumar V, Chan AP, Thibaud-Nissen F, Schobel S, Town CD (2017) Araport11: a complete reannotation of the *Arabidopsis thaliana* reference genome. Plant J 89:789–804

Chumak N, Mosiolek M, Schoft VK (2015) Sample preparation and fractionation of *Arabidopsis thaliana* sperm and vegetative cell nuclei by FACS. Bio-protocol 5:e1664

Dupl'áková N, Dobrev PI, Reňák D, Honys D (2016) Rapid separation of Arabidopsis male gametophyte developmental stages using a Percoll gradient. Nat Protoc 11:1817

Efroni I, Birnbaum KD (2016) The potential of single-cell profiling in plants. Genome Biol 17:65

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

Friedman WE (1999) Expression of the cell cycle in sperm of Arabidopsis: implications for understanding patterns of gametogenesis and fertilization in plants and other eukaryotes. Development 126:1065–1075

Gawad C, Koh W, Quake SR (2016) Single-cell genome sequencing: current state of the science. Nat Rev Genet 17:175

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

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

Ibarra CA, Feng X, Schoft VK, Hsieh TF, Uzawa R, Rodrigues JA, Zemach A, Chumak N, Machlicova A, Nishimura T, Rojas D (2012) Active DNA demethylation in plant companion cells reinforces transposon methylation in gametes. Science 337:1360–1364

Jin X, Wang RS, Zhu M, Jeon BW, Albert R, Chen S, Assmann SM (2013) Abscisic acid–responsive guard cell metabolomes of Arabidopsis wild-type and gpa1 G-protein mutants. Plant Cell 25:4789–4811

Kim D, Langmead B, Salzberg SL (2015) HISAT: a fast spliced aligner with low memory requirements. Nat Methods 12:357–360

Leydon AR, Weinreb C, Venable E, Reinders A, Ward JM, Johnson MA (2017) The molecular dialog between flowering plant reproductive partners defined by SNP-informed RNA-Sequencing. Plant Cell 29:984–1006

Libault M, Brechenmacher L, Cheng J, Xu D, Stacey G (2010) Root hair systems biology. Trends Plant Sci 15:641–650

Libault M, Pingault L, Zogli P, Schiefelbein J (2017) Plant systems biology at the single-cell level. Trends Plant Sci 22:949–960

Lin S-Y, Chen PW, Chuang MH, Juntawong P, Bailey-Serres J, Jauh GY (2014) Profiling of translatomes of in vivo-grown pollen tubes reveals genes with roles in micropylar guidance during pollination in Arabidopsis. Plant Cell 26:602–618

Liu L, Lu Y, Wei L, Yu H, Cao Y, Li Y, Yang N, Song Y, Liang C, Wang T (2018) Transcriptomics analyses reveal the molecular roadmap and long non-coding RNA landscape of sperm cell lineage development. Plant J 96:421–437

Macaulay IC, Teng MJ, Haerty W, Kumar P, Ponting CP, Voet T (2016) Separation and parallel sequencing of the genomes and transcriptomes of single cells using G&T-seq. Nat Protoc 11:2081–2103

Macaulay IC, Ponting CP, Voet T (2017) Single-cell multiomics: multiple measurements from single cells. Trends Genet 33:155–168

Mantsoki A, Devailly G, Joshi A (2016) Gene expression variability in mammalian embryonic stem cells using single cell RNA-seq data. Comput Biol Chem 63:52–61

Marinov GK, Williams BA, McCue K, Schroth GP, Gertz J, Myers RM, Wold BJ (2014) From single-cell to cell-pool transcriptomes: stochasticity in gene expression and RNA splicing. Genome Res 24:496–510

Marks MD, Betancur L, Gilding E, Chen F, Bauer S, Wenger JP, Dixon RA, Haigler CH (2008) A new method for isolating large quantities of Arabidopsis trichomes for transcriptome, cell wall and other types of analyses. Plant J 56:483–492

Martin M (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet J 17:10–12

Mogensen HL (1992) The male germ unit: concept, composition, and significance. Int Rev Cytol 140:129–147

Okonechnikov K, Conesa A, García-Alcalde F (2015) Qualimap 2: advanced multi-sample quality control for high-throughput sequencing data. Bioinformatics 32:292–294

Palanivelu R, Preuss D (2006) Distinct short-range ovule signals attract or repel *Arabidopsis thaliana* pollen tubes in vitro. BMC Plant Biol 6:7

Patro R, Duggal G, Love MI, Irizarry RA, Kingsford C (2017) Salmon provides fast and bias-aware quantification of transcript expression. Nat Methods 14:417–419

Picelli S, Faridani OR, Björklund ÅK, Winberg G, Sagasser S, Sandberg R (2014) Full-length RNA-seq from single cells using Smart-seq2. Nat Protoc 9:171

Pina C, Pinto F, Feijó JA, Becker JD (2005) Gene family analysis of the Arabidopsis pollen transcriptome reveals biological implications for cell growth, division control, and gene expression regulation. Plant Physiol 138:744–756

Puram SV, Tirosh I, Parikh AS, Patel AP, Yizhak K, Gillespie S, Rodman C, Luo CL, Mroz EA, Emerick KS, Deschler DG (2017) Single-cell transcriptomic analysis of primary and metastatic tumor ecosystems in head and neck cancer. Cell 171:1611–1624

Qiao Z, Libault M (2013) Unleashing the potential of the root hair cell as a single plant cell type model in root systems biology. Front Plant Sci 4:484

Qin Y, Leydon AR, Manziello A, Pandey R, Mount D, Denic S, Vasic B, Johnson MA, Palanivelu R (2009) Penetration of the stigma and style elicits a novel transcriptome in pollen tubes, pointing to genes critical for growth in a pistil. PLoS Genet 5:e1000621

Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, Mesirov JP (2011) Integrative genomics viewer. Nat Biotechnol 29:24–26

Russell SD (1985) Preferential fertilization in Plumbago: ultrastructural evidence for gamete-level recognition in an angiosperm. Proc Natl Acad Sci USA 82:6129–6132

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

Saito C, Nagata N, Sakai A, Mori K, Kuroiwa H, Kuroiwa T (2002) Angiosperm species that produce sperm cell pairs or generative cells with polarized distribution of DNA-containing organelles. Sex Plant Reprod 15:167–178

Santos MR, Bispo C, Becker JD (2017) Isolation of arabidopsis pollen, sperm cells, and vegetative nuclei by fluorescence-activated cell sorting (FACS). In: Schmidt A (ed) Plant germline development. Humana Press, New York, pp 193–210

Slotkin RK, Vaughn M, Borges F, Tanurdžić M, Becker JD, Feijó JA, Martienssen RA (2009) Epigenetic reprogramming and small RNA silencing of transposable elements in pollen. Cell 136:461–472

Twell D (2011) Male gametogenesis and germline specification in flowering plants. Sex Plant Reprod 24:149–160

Villani AC, Satija R, Reynolds G, Sarkizova S, Shekhar K, Fletcher J, Griesbeck M, Butler A, Zheng S, Lazo S, Jardine L (2017) Single-cell RNA-seq reveals new types of human blood dendritic cells, monocytes, and progenitors. Science 356:eaah4573

Wang L, Wang S, Li W (2012) RSeQC: quality control of RNA-seq experiments. Bioinformatics 28:2184–2185

Yang C, Ye Z (2013) Trichomes as models for studying plant cell differentiation. Cell Mol Life Sci 70:1937–1948

Zhang J, Huang Q, Zhong S, Bleckmann A, Huang J, Guo X, Lin Q, Gu H, Dong J, Dresselhaus T, Qu LJ (2017) Sperm cells are passive cargo of the pollen tube in plant fertilization. Nat Plants 3:17079