The Role of Tapetal Nurse Cells in Supporting Male Germline Functions

Billy Aldridge

John Innes Centre

Thesis for the degree of *Doctor of Philosophy* (*PhD*)

Submitted for examination to the University of East Anglia

September 2018

Abstract

The development of germlines is an important timepoint in the lifecycles of multicellular organisms. In plants a somatic cell layer, called the tapetum, supports the development of the male germline. The tapetum controls many aspects of germline development, and as such has been an important target for research on male fertility. While large effect genes have been identified in screens for male sterility, tapetal research is hampered by the difficulty in isolating high-purity tapetum and the lack of whole genome sequencing data. In this thesis, tapetum function was explored by the application of fluorescence activated cell sorting to *Arabidopsis thaliana* plants expressing a tapetal GFP reporter.

RNA-sequencing of isolated tapetal cells has led to the discovery of novel tapetum-enriched and tapetum-specific genes. Mutant plants of these genes show defects in both tapetum and pollen development. Sequencing of single tapetal cells has allowed the inference of gene expression through developmental time. Tapetum gene expression could be classified into two broad patterns, early and late, with a developmental switch inferred at anther stage 8. Analysis of temporal gene expression has revealed new gene expression profiles of known tapetal genes as well as novel candidate genes expressed at specific developmental stages. Whole genome DNA methylation analysis of the tapetum has revealed sites of tapetum-specific hypermethylation controlled by both canonical and non-canonical RNA-directed DNA methylation pathways. The observed methylation patterns in the tapetum and sexual lineage has led to a model of small RNA transfer from the tapetum.

Contents

Acknowledgements

Throughout the process of writing this thesis I've often looked back on how I got here and have concluded the blame should rest squarely at the feet of my parents. Mum, Dad, thank you for always encouraging me to do what I love. I owe much to my supervisors at the Cambridge plants department, whose passion and drive were infectious. Howard Griffiths was an enthusiastic and encouraging mentor throughout, who taught me so much, including the most important lesson from Cambridge: you can't drink singles!

The rotation programme has allowed me to explore the full breadth of science at the John Innes Centre and gave me the skills needed to tackle a project I could really call my own. Thanks to Nick Brewin and Steph Bornemann for creating a wonderful programme, and to the John Innes Foundation for funding this research throughout.

To my supervisor, Xiaoqi Feng: when I first arrived for my rotation it was just you and me in an empty lab where the most exciting piece of equipment was a hairdryer. Now, I'm finishing a PhD as part of one of the most dynamic, welcoming, and exciting labs at the JIC. It has been a privilege to be here from the very start and see the lab grow into what it is today, thank you.

This project has been supported throughout by great collaborations. Stefan Scholten an ever-reliable collaborator who created the sequencing libraries on which this whole project has depended. Working with Iain Macaulay has enabled the exploration of single-cell sequencing, which has been one of the most fun projects I've worked on. Thanks too to my supervisory committee, Lars Østergaard and Hongbo Gao, for their insight, helping steer this project through to the end.

I have been incredibly fortunate to be able to rely on the help of brilliant postdocs; Hongbo Gao, Martin Vickers, Shengbo He, and Jingyi Zhang. Martin has taught me practically everything I know about bioinformatics and weathered my stupid questions with the patience of a saint. Jingyi has helped with so much of this project and done it all with a warmth and generosity which is unrivalled.

Thank you to my fellow students; James, Toby, and Sam. You've all gone above and beyond helping me through the last stage of my PhD. I have been so lucky to work with such great students and amazing scientists.

The biggest thanks go to my friends and family who have helped me through the many highs and lows. Thanks for your support, care, sympathy, and patience. Megaladz, thanks for keeping me (semi)sane these past months.

And finally, Dan, thanks for the snacks, the dog videos, and your unconditional support. It's impossible to sum up how much you've helped me but it's fair to say, I wouldn't have made it this far without you.

Work by others

Throughout this thesis my work is built on work by others in the lab and performed by collaborators. Stefan Scholten (university of Hohenheim) create the WT tapetal RNA-seq libraries and BS-seq libraries from WT, *rdr2*, and *rdr6*. Iain Macaulay and Ashleigh Lister (Earlham Institute) created the single-cell mRNA sequencing libraries. mRNA read mapping was performed by Graham Etherington (Earlham Institute), while remaining single-cell analyses were performed by me. Jingyi Zhang (Feng lab) performed the majority of resin sectioning and all Alexander staining. SEM imaging was performed by Elaine Barclay (JIC; Bioimaging). James Walker and Samuel Deans (Feng lab) performed sRNA mapping and analysis. Hongbo Gao (Feng lab) extracted meiocytes, created *rdr2* tapetal mRNA-seq libraries and performed immunostaining.

Chapter 1 Introduction

All sexually reproducing multicellular organisms develop from the fusion of just two cells, the male and female gametes. These cells represent the germline, the cell lineage that will propagate the genome through generations. How germlines develop and are maintained is a central question to the study of species and their evolution. Germlines do not form on their own and are dependent on somatic tissues for their development. In plants, the male germline forms within pollen grains. This development is supported by a somatic cell layer called the tapetum. The tapetum is essential for the production of pollen and performs a variety of functions to support pollen development. While tapetum function has been studied in both model and crop species, research has been hindered by the difficulty in isolating tapetal cells, and the lack of whole genome sequencing data available.

In this thesis, I have investigated gene expression and DNA methylation in the tapetum of *Arabidopsis thaliana*. I will start by introducing the development of the *Arabidopsis* tapetum, and the known roles it plays in reproductive development. I will then briefly introduce DNA methylation and chromatin modification pathways, before discussing their role in reproductive development and how they relate to the results presented in the thesis.

1.1 The tapetum and its roles in supporting male germline development

Sexual reproduction is a crucial timepoint in the life cycle of many organisms. Development must be coordinated to produce healthy gametes at the right time and in the right number. The development of pollen occurs surrounded by a somatic cell layer, the tapetum. The tapetum is present throughout the development of the male sexual lineage and performs a wide range of functions, which are summarised below.

1.1.1 Anther development

The plant life cycle can be divided into two distinct phases of haploid and diploid growth. The diploid phase is called the sporophyte because it produces haploid spores by meiosis. These haploid spores form the gametophyte generation, which produce male and female gametes by mitosis [1]. The development of these generations varies greatly, with different plant lineages having different dominant stages. Liverworts are gametophyte-dominant. Ferns are sporophyte-dominant, but with a free-living gametophyte stage. In angiosperms (flowering plants), the sporophyte is dominant, and the gametophyte is reduced to a few mitotic divisions, wholly dependent on the mature sporophyte.

Most angiosperms produce male and female gametophytes within the same flower; the male gametophyte develops from a microspore, and the female from a megaspore [1]. Floral organs are arranged in four concentric whorls, from the outer- to innermost; sepals, petals, stamen, and carpels. In *Arabidopsis thaliana,* flowers consist of four sepals and petals, six stamens, and a single carpel. Specific floral organs are defined by the combinatorial action of homeotic genes, which are mostly MADS-box transcription factors [2]. A-function genes act in sepals and petals. B-function genes act in petals and stamens. C-function genes act in stamens and carpels [3]. As well as the action of the ABC genes, the development of petals, stamens, and carpels require the action of E-function, '*SEPELLATA*', genes. Ectopic expression of E-function genes with combinations of A-, B-, and C-function genes can convert leaves into various floral organs [4, 5].

After definition of male and female floral organs, gametophyte development can occur. As the germline is the immortal lineage that transmits the genome to the next generation, there is an evolutionary pressure to protect germlines from mutations. This highlights a difference between animal and plant germline development. In many animals, the germline cells are defined early in development and sequestered (guarding against mutation). Conversely, plants define their germlines from somatic cells after the switch from vegetative to reproductive development [6]. It has been

argued, however, that plants do functionally segregate their germlines through the slow cell division of meristems [7], and subsequently differentiate their germlines late in development.

Male germlines develop in the stamens of flowers. The stamen consists of a long filament supporting the anther. The anther is a four-lobed structure containing the male sexual lineage. At the centre of each lobe, the meiocytes (also called microsporocytes) develop in a process known as sporogenesis. Meiocytes produce the male gametophytes (Fig. 1-1) [8]. Once the meiocytes have progressed through meiosis, the four meiotic products develop into microspores (Fig. 1-1). The process of producing gametes from haploid spores is known as gametogenesis. Microspores mature in the anther locule, going through one round of mitosis to give rise to the vegetative cell encasing the generative cell. The generative cell then performs another round of mitosis to give rise to the true germline (gametes): the two sperm cells (Fig. 1-1). Development from the microspores to the final pollen mitosis represents the full extent of gametophyte development in angiosperms. While meiocytes are not the true germline, they give rise to cells which develop into the germline. In keeping with Walker & Gao et al., 2018 [9], throughout this thesis, the lineage from meiocytes to mature pollen will be referred to as the 'sexual lineage' and 'germline' will refer to the sperm cells and the female egg cell (Fig. 1-1). Germline development broadly represents processes which affect the production of viable pollen and sperm cells.

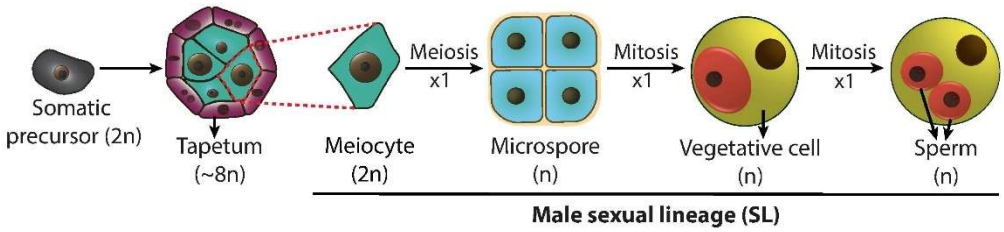


Figure 1-1: The male sexual lineage in *Arabidopsis thaliana*. The sexual lineage, and the surrounding somatic cell layers, develop from somatic precursor cells in the anther. The meiocytes are surrounded by the tapetum, which supports the development of the sexual lineage. Meiocytes undergo meiosis to give rise to four haploid microspores. Microspores will then divide by mitosis to produce a vegetative cell surrounding a smaller generative cell inside a pollen grain. The generative cell then divides by mitosis again to produce two sperm cells. These sperm cells are the true germline, as they will pass on their genomes during sexual reproduction. Figure reproduced with permission from Hongbo Gao.

1.1.2 Anther developmental staging

In *Arabidopsis thaliana,* the development of anthers has been studied in detail with the use of male-sterile mutants and transverse sectioning [10, 11]. The progression of stages appears almost identical in other angiosperms, such as rice, highlighting the high degree of conservation in anther development [12]. At flower stage 5, the rounded stamen primordia emerge in the developing flower, and so this is defined as anther stage 1. At anther stage 1 the primordia consist of the L1, L2, and L3 layers [13] that will develop into the various anther tissues. The archesporial cells (which develop into meiocytes) and epidermis are defined by anther stage 2. The secondary parietal cells (which give rise to the middle layer and tapetum) arise at anther stage 3.

By stage 4, sporogenesis and the development of the somatic cell layers is complete, and the anther begins to take on its characteristic shape [10, 14]. At stage 5, clearly defined locules are established with the meiocytes present and connected to the tapetum via plasmodesmata (Fig. 1-2) [15]. At stage 6, the meiocytes enter meiosis and are surrounded by a callose cell wall (Fig. 1-2). The tapetum undergoes endoreplication: DNA replication without cell division [16]. Meiosis is completed by stage 7 and tetrads of microspores are free within each locule (Fig. 1-2). The tapetum will then produce enzymes to degrade the callose wall surrounding the tetrads, releasing individual microspores at stage 8 (Fig. 1-2) [17].

Stage 9 sees the rapid production of the pollen wall and at this stage the tapetum is highly active in secreting sporopollenin precursors (Fig. 1-2). With the pollen wall produced, the tapetum will then enter programmed cell death (PCD) at stage 10 [18]. This is complete by stage 11, where tapetum degeneration will release material onto the pollen grains, particularly fatty acids (Fig. 1-2). This material forms the pollen coat. Stage 11 also sees pollen mitosis I occur, giving rise to the generative cell encompassed by the vegetative cell. Pollen mitosis II is completed by stage 12 (in *Arabidopsis thaliana*) where the anther becomes bilocular after the degeneration of the

septum between each pair of locules (Fig. 1-2). Anther dehiscence occurs at stage 13, releasing the pollen. At stage 14, the whole stamen senesces.

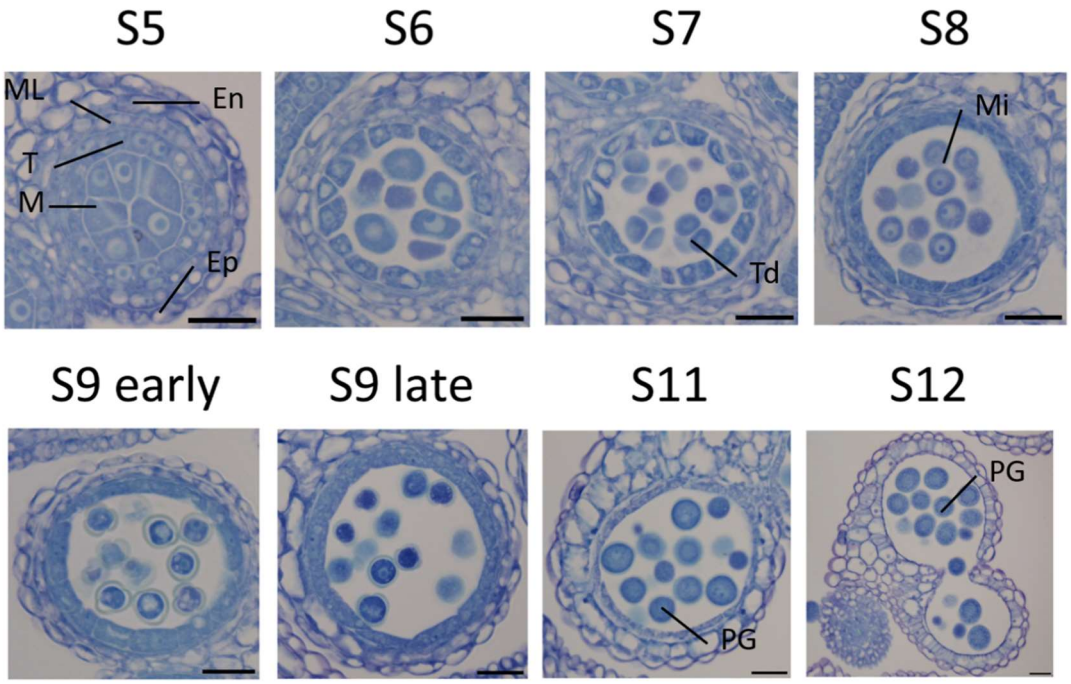


Figure 1-2: Anther stages from Arabidopsis thaliana inflorescence cross-sections. At stage 5 all anther cell types have developed and can be seen in cross-section. The meiocytes (M) are surrounded by four somatic cell layers, the tapetum (T), middle layer (ML), endothecium (En), and epidermis (Ep). At stage 6 the meiocytes are surrounded by a callose cell wall and undergo meiosis. The middle layer becomes crushed. By stage 7 meiosis is complete and tetrads of microspores are free within the anther locule. The callose wall surrounding the tetrads is broken down at stage 8, releasing individual microspores (Mi) into the locule. The pollen exine wall is generated at stage 9 and the microspores become vacuolated, though they are less so by late stage 9. Tapetum degeneration is initiated at stage 10 and can be seen at stage 11. Pollen mitosis also occurs at stage 11, producing pollen grains (PG). By stage 12 tapetum degeneration is complete and anthers become bilocular after the degeneration of the septum dividing locules. Anther dehiscence occurs at stage 13 (not pictured), releasing pollen grains. Stages are taken from [10] and [11]. Scale bars represent 20 μm.

1.1.3 Tapetum development

Surrounding the sexual lineage is a series of somatic cell layers arranged in concentric rings. Closest to the sexual lineage is the tapetum, then the middle layer, the endothecium, and finally the epidermis [10]. The anther primordium that emerges from the floral meristem consists of three cell layers L1-3 [13]. The L1 layer forms the epidermis (the outside of the anther). L3 forms the connective tissue. L2 forms the sexual lineage and surrounding somatic cell layers. In what way somatic and sexual lineage cell layers form from L2-derived (L2-d) cells has been debated for some time.

A lineage model had been proposed, wherein archesporial cells develop from L2-d cells and then give rise to the sporogeneous cells (cells which produce meiocytes), tapetum, middle layer and endothecium. Each cell layer was proposed to be defined by asymmetric cell divisions of the parent cell [19]. Evidence from maize cast doubt on this model, because it was seen that archesporial cells develop early and the somatic cell layers develop from other L2-d cells (Fig. 1-3) [20]. It was shown that L1-d cells can differentiate into archesporial cells under low Oxygen conditions, suggesting cell fate is determined positionally [20]. Expression of tapetum development genes can be used as markers of early stage cells. Next, some of the mechanisms which act in the anther to determine cell fate will be discussed.

Hypoxia as a positional cue

Cellular redox status is an important factor in male gametogenesis across the angiosperms. In *Arabidopsis thaliana,* two glutaredoxins (ROXY1 and ROXY2) act redundantly in early anther development [21]. *ROXY1* and *ROXY2* expression is detectable in anther lobe primordia when the archesporial cells differentiate into sporogeneous cells. They are also strongly expressed in meiocytes and somatic cell layers prior to meiosis. The *roxy1;roxy2* double mutant displays male sterility with early defects including incorrect abaxial–adaxial anther lobe formation and a failure to differentiate meiocytes [21]. The maize ROXY homologue MALE STERILE CONVERTED ANTHER1 (MSCA1), and the rice MICROSPORELESS 1 (MIL1) also function in cell fate determination; precursor cells in each mutant fail to differentiate into mature archesporial cells [22, 23]. It was also found that varying oxygen levels can affect germ cell number [20], suggesting that hypoxia, occurring naturally in growing anthers, serves an evolutionarily conserved role as a positional cue to define meiocyte fate [14, 24] (Fig. 1-3).

Peptide signalling

While hypoxia is essential for the definition of archesporial cells (and therefore meiocytes), the somatic cell layers are defined by other means. In *Arabidopsis,* two

CLAVATA1-like LRR-RLKs, BARELY ANY MERISTEM 1 (BAM1) and BAM2 function partially redundantly to define the endothecium, middle layer and tapetum by limiting the expression of *SPOROCYTELESS/NOZZLE* (*SPL/NZZ*). SPL/NZZ is a MADS-box transcription factor required for promoting the formation of parietal somatic and reproductive cells (Fig. 1-3) [25, 26].

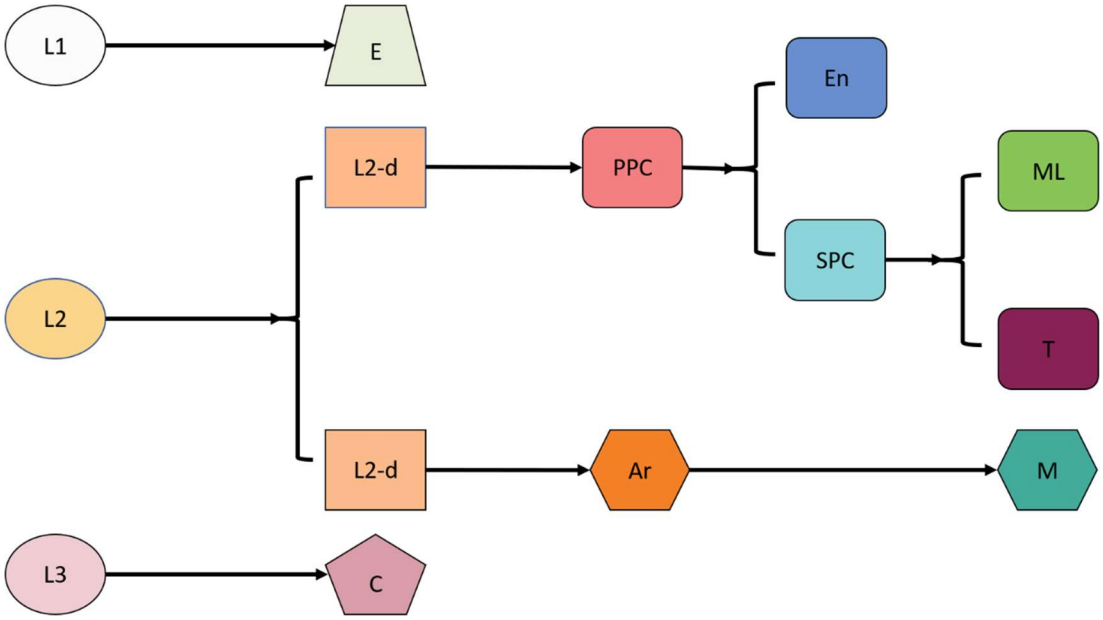


Figure 1-3: Diagram of cell lineages during early anther development. The anther primordium has three cell layers; the L1, L2 and L3 layers: L1 differentiates into the epidermis (E); L2 cells generate the endothecium (En), middle layer (ML), tapetum (T) and meiocytes (M); L3 forms the connective tissues Ar: archesporial cell; C: connective tissue; L2-d: L2-derived cell; PPC: primary parietal cell ; SPC: secondary parietal cell. Figure adapted with permission from [14].

A further LRR-RLK complex also functions in the definition of the tapetum, consisting of EXCESS MICROSPOROCYTES 1 (EMS1) and SOMATIC EMBRYOGENESIS RECEPTOR LIKE KINASE 1 (SERK1) or SERK2 [27-30].The EMS1 receptor is expressed on the cell surface of tapetal precursor cells and the ligand TAPETAL DETERMINANT 1 (TPD1) is excreted from the sporogeneous cells to promote tapetal fate [31-33]. Mutants such as e*ms1, serk1;serk2* and *tpd1* lack tapetal cells and have excess meiocytes that cannot complete meiosis. Exogenous expression of EMS1 in an *ems1* mutant can rescue tapetal cell fate, suggesting that anthers contain early-defined meiocyte and tapetal cells and the excess meiocytes seen in *ems1* are due to over-proliferation, not an increase in the number of cells defined as meiocytes [34]. Together hypoxia and peptide signalling pathways act to tightly control the

position and number of cells in the anthers to produce functional pollen. Once defined, the tapetum then performs a range of functions to support the sexual lineage.

1.1.4 Endoreplication

Endoreplication is a cell cycle variant in which cells replicate DNA without proceeding to mitosis. This produces cells that are polyploid and/or multinucleate [35]. Endoreplication often occurs in metabolically active or highly specialised cells [35]. In many species, the tapetum is endoreplicated. *Arabidopsis thaliana* tapetal cells are polyploid and multinucleate, most often possessing two tetraploid nuclei [16]. However, the end state of tapetum endoreplication is not fixed and cells of different ploidy level and nuclei number are observed at the same stages [16]. Tapetal nuclei division in spinach (*Spinacia oleracea*) can give rise to binucleate cells, cells with lobed polyploid nuclei, or cells where chromosome replication occurs without nuclear division [36]. In some *Phaseolus* species (*Fabaceae*), tapetal nuclei possess polytene chromosomes where sister chromatids are fused together [37]. Locus-specific endoreplication occurs in *Phaseolus coccineus* suspensor cells, where satellite DNA is replicated to a higher level than ribosomal RNA genes [38]. *Arabidopsis* tapetum endoreplication level was inferred from the level of 45S ribosomal genes [16], and it is unknown if locus-specific endoreplication occurs.

While endoreplication has been proposed to enable a greater rate of metabolism [35, 39], the ultimate reason for endoreplication in the tapetum is not known. Endoreplication is promoted by the inhibition of cyclin-dependent kinase activity below the threshold required for mitosis. This inhibition occurs at both the transcriptional and post-transcriptional level [35]. While progress has been made in elucidating genes involved in the regulation of endoreplication [35], those affecting tapetal endoreplication remain unknown.

1.1.5 Pollen wall biosynthesis

For most plant species, pollen is dispersed long distances by wind or animal vectors before reaching the stigma of another flower. As such, pollen grains are well protected against the external environment by a complex and highly-resistant cell wall. The shape and structure of this wall varies hugely between plants and can be used to identify different species, even from ancient samples. In *Arabidopsis thaliana,* pollen grains are long and oval shaped, divided into three lobes by invaginations of the pollen wall along its length, called apertures (Fig. 1-4). The outer surface of the pollen wall forms a reticulate pattern, with cavities in this structure filled with pollen coat material (Fig. 1-4).

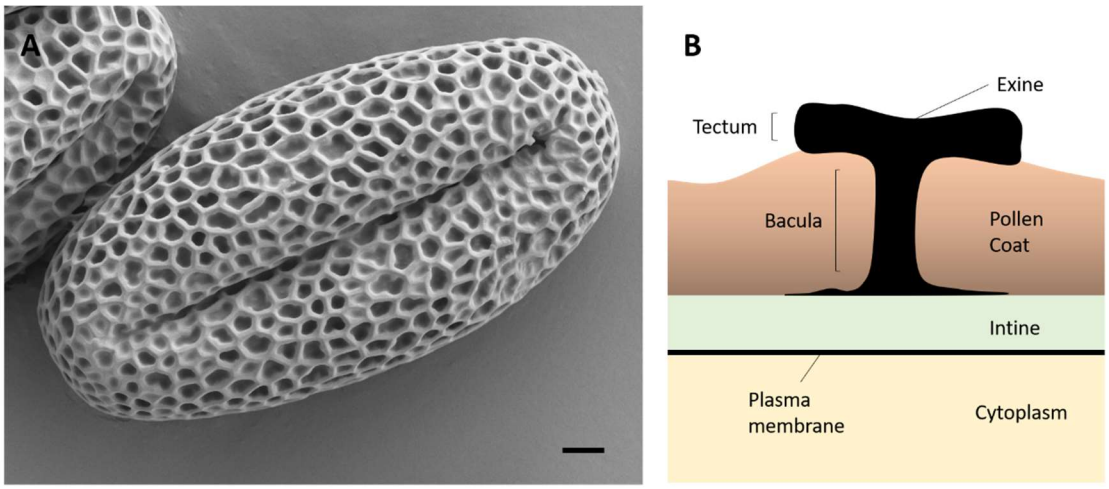


Figure 1-4: Overview of pollen wall structure. **A**) SEM image of a Col-0 WT Arabidopsis thaliana pollen grain. Pollen are long and oval, divided into three lobes by invaginations of the pollen wall called apertures. Scale bar represents 2µm. **B**) A schematic of a pollen wall cross section. The lighter-staining intine sits above the plasma membrane and is composed of cellulose, hemicellulose and pectin. The exine is made of sporopollenin and is produced by the tapetum. The exine consists of pillars (baculae) supporting large caps of sporopollenin (tecta). The exine forms the reticulate pattern of the pollen wall seen in A. Between the baculae and tecta, cavities become filled with pollen coat material, consisting largely of lipids.

The pollen wall is a complex of multiple layers formed by both the gametophyte (microspore) and the sporophyte (tapetum). The inner pollen wall, the intine, is produced by the microspore and is composed of cellulose, pectin, and hemicellulose [40]. The outer wall, the exine, is produced by the tapetum and is formed of sporopollenin [40]. Sporopollenin is a complex biopolymer consisting of long chain fatty acids, phenylpropanoids, phenolics and traces of carotenoids. These molecules

become cross-linked by ethyl and ester linkages on the surface of the developing microspores [41]. The exine forms the complex structure visible on the pollen surface, consisting of pillars of sporopollenin, called baculae. These pillars support large caps, 'tecta', which connect in the reticulate pattern seen in *Arabidopsis* (Fig.1-4). Pollen coat material (also called tryphine), largely consisting of lipids and proteins, is deposited between the baculae and tecta. Pollen coat material aids stigmatic adhesion and pollen recognition in self-incompatible species [41-43]. The creation of this complex cell wall requires coordination between the tapetum and gametophyte in their development and metabolism. Below, some of the processes and genes identified in regulating *Arabidopsis* pollen wall formation will be summarised, with a focus on the role played by the tapetum.

While the details of pollen wall development are not essential knowledge for this thesis, it is important to know that many genes involved in sporopollenin biosynthesis are important stage markers and are key targets of tapetum-expressed transcription factors. The mis-regulation of these genes often leads to male sterility.

Microspore primary cell walls

At the early meiosis stage (stage 6), callose is deposited beneath the pectin cell wall of the meiocytes and forms the major component of the tetrad cell wall [44]. Callose metabolism is regulated by *CALLOSE DEFECTIVE MICROSPORE 1* (*CDM1*) which is expressed in both the meiocytes and tapetum [45]. c*dm1* plants are male sterile due to collapse of microspores, showing the importance of callose deposition for male germline development [45]. The callose cell wall is essential for the development of exine structure on the pollen surface. At the microspore stage, the plasma membrane undulates as a primexine cell wall is deposited by the microspore [40]. It is proposed that the callose cell wall forms a hard barrier around the microspore, forcing the plasma membrane to undulate beneath the soft matrix of the primexine [46]. This plasma membrane undulation then forms the pattern upon which the exine is built.

Callose and pectin must be degraded for individual microspores to be released from tetrads. In *Arabidopsis* and *Brassica* species, the β-1,3-glucanse protein A6 is expressed in the tapetum and is predicted to degrade the callose cell wall surrounding the tetrads [17]. Pectin is degraded by the tapetum-expressed QUARTET (QRT) proteins [44]. In *qrt* mutants, callose is successfully degraded, but microspores remain connected by the pectin remnants of the meiocyte cell wall [44]. Beneath the callose cell wall of the tetrads, sporopollenin deposition begins along the pattern laid out by the primexine [46]. Sporopollenin is a highly complex material, and as such many pathways act to produce precursors in the tapetum.

Sporopollenin biosynthesis

Lipids form a major constituent of exine and, in the tapetum, sporopollenin biosynthesis is closely associated with fatty acid metabolism. The tapetal plastids form specialised organelles called elaioplasts, where fatty acid synthesis occurs [47]. The enzyme ACYL-CoA SYNTHETASE 5 (ACOS5) acts on medium to long chain fatty acids to create acyl-CoA esters, which are acted upon by further sporopollenin biosynthesis enzymes [48]. Highlighting this early role in sporopollenin biosynthesis, *acos5* mutant plants are completely male sterile [48]. The chalcone synthase-like enzymes LESS ADHERENT POLLEN 5 (LAP5) and LAP6 catalyse the condensation of two or three malonyl-CoAs, with ACOS5-produced acyl-CoAs, to produce tri- and tetraketide α-pyrones [49] (Fig. 1-5c). The tetraketide products of LAP5 and LAP6 are substrates for TETRAKETIDE α-PYRONE REDUCTASE 1 (TKPR1) and TKPR2 [41, 50]. The reduced hydroxy-tetraketide α-pyrones, produced by TKPR1&2, possess an extra hydroxyl group on the acyl chain. Hydroxyl groups act as an additional site for the formation of ester and ether linkages in polymerised sporopollenin (Fig. 1-5c) [42]. Plants with a mutant allele of *TKPR1* show a strong male-sterile phenotype and a severely disturbed exine [50]. Hydroxyl groups are also added by the action of Cytochrome p450 enzymes, CYP704A2 [51] and CYP703B1 [52]. Both enzymes act on the fatty acid constituents of sporopollenin, with differing specificities [41]. These

enzymes act together in the endoplasmic reticulum (ER) to produce hydroxylated tri- and tetraketide molecules for sporopollenin polymerisation (Fig. 1-5c).

Very long chain fatty acids (VLCFAs) and waxes are important sporopollenin/pollen coat components. VLCFA-derived fatty alcohols have been shown to form a key subset of sporopollenin constituents (Fig. 1-5a). *MALE STERILITY 2* (*MS2*) was identified as a gene required for exine formation, mutants of which show severe male fertility defects [53]. MS2 functions as a plastid-localised fatty acyl reductase to produce fatty alcohols [53]. VLCFAs are also converted to wax components by decarbonylation, yielding aldehydes, alkanes, secondary alcohols and ketones [42]. Two genes have been implicated in the production of these molecules in the tapetum: *ECERIFERUM 1* (*CER1*) and *CER3/FACELESS POLLEN 1* (*FLP1*) [54-56].

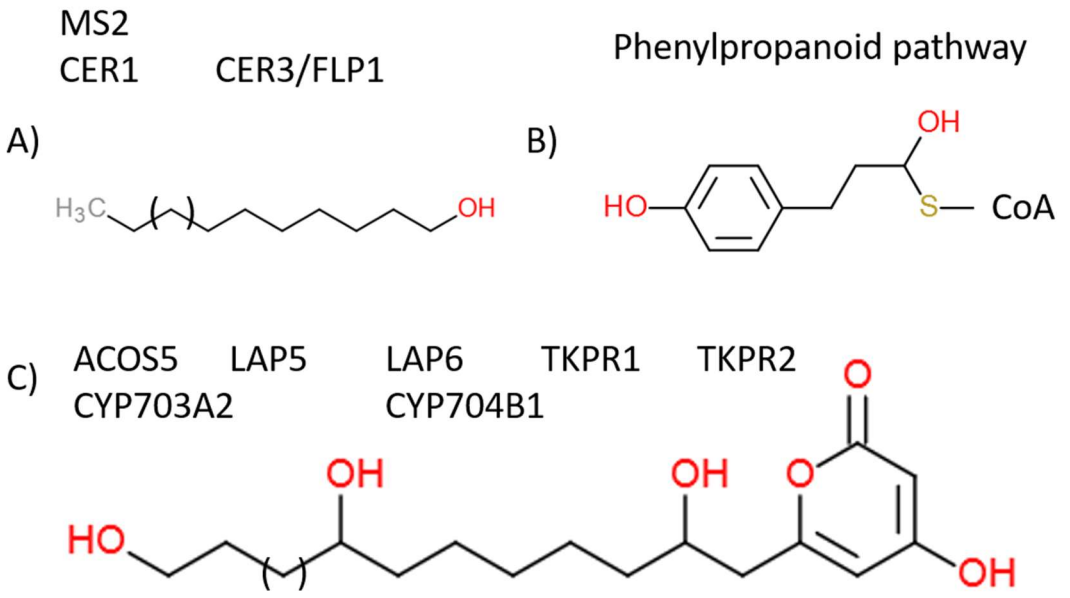


Figure 1-5: **A**) An example of a fatty alcohol, produced by MS2. CER1 and CER3/FLP1 act to produce similar alkanes, ketones, aldehydes, and secondary alcohols. **B**) p-Coumaroyl CoA, an example phenolic compound produced in the phenylpropanoid pathway. **C**) An example of a hydroxylated tetraketide α-pyrone, produced by the action of ACOS5, LAP5, LAP6, TKPR1, TKPR2, CYP703A2 and CYP704B1.

As well as fatty acids, phenolic compounds are an important component of sporopollenin (Fig. 1-5b) [57]. As in the production of lignin, phenolics are derived from phenylalanine, via phenylalanine ammonia lyases [58]. Sporopollenin phenolic compounds appear to be derived from hydroxycinnamic acids of the phenylpropanoid pathway. A mutant of the *CINNAMATE-4-HYDROXYLASE* gene, which encodes a key entry point enzyme into phenylpropanoid metabolism, fails to develop pollen and the pollen wall lacks fluorescence (derived from phenolic compounds) [59]. On the surface of the pollen, phenolics and hydroxylated-tetraketides can crosslink, forming an irregular but rigid structure [57]. The phenylpropanoid pathway is also required to produce flavonoids in the tapetum, along with the gene *LAP3* [42, 60].

Transport of pollen wall and pollen coat material

Sporopollenin precursors are assembled on the surface of the microspore following the guide of the primexine, suggesting that sporopollenin precursors are actively exported from the tapetum. The *ATP-BINDING CASSETTE G26* (*ABCG26*) gene encoding an ABC half transporter protein has been proposed to export sporopollenin precursors, analogous to wax export from epidermal cells [61]. Lipidic components of the exine are exported bound to lipid transfer proteins (LTPs), many of which have been identified as anther- or tapetum-specific [62, 63]. In *Arabidopsis*, ARABIDOPSIS THALIANA ANTHER7 (ATA7/A7)[64] and A9 [65] are tapetum-expressed lipid transfer proteins.

Unlike sporopollenin precursors, the material that forms the pollen coat builds up in the tapetum and is released onto the pollen surface upon tapetal PCD (reviewed in section 1.1.6). These fatty acid components are synthesised and stored in specialised organelles; tapetosomes and elaioplasts. Tapetosomes are ER-derived storage organelles that produce a wide range of metabolites, including alkanes and triacylglycerol rich oil droplets. Oil bodies are structurally maintained by oleosins and are also surrounded by vesicles associated with flavonoids [66-68]. When viewed by transmission electron microscopy, tapetosomes are electron dense structures

associated with the ER [11]. Elaioplasts are rich in numerous steryl esters, free polar lipids and plastid lipid-associated proteins [41, 47]. The tapetosomes and elaioplasts increase in size and number late in tapetal development and come to occupy the majority of the cellular space [11].

1.1.6 Programmed cell death

After the production of the exine pollen wall the tapetum will undergo PCD, releasing cell contents onto the pollen grains. While the study of PCD is relatively mature in animals, much less is understood about PCD in plant systems [69]. The male sterility defects associated with precocious or delayed tapetal PCD have facilitated the identification of PCD-related genes [18]. Phytohormone signalling, particularly of gibberellic acid (GA), has been shown to function at the onset of tapetum PCD. In rice, the GA-responsive GAMYB transcription factor regulates both exine formation and PCD [70]. The *Arabidopsis* GAMYB-homologues MYB33 and MYB65 regulate tapetal development [71, 72]. *Arabidopsis* CYSTEINE ENDOPEPTIDASE 1 (CEP1) has been shown to function in tapetum PCD [73]. CEP1 is expressed from stage 5, long before visible tapetum degeneration, as a proenzyme in precursor protease vesicles. CEP1 then matures in the vacuole and acts in PCD after the rupture of the vacuolar membrane. Both CEP1 depletion and overexpression severely reduce male fertility, highlighting the importance of PCD timing for male germline development.

Negative regulators of PCD have also been discovered. For example, plants with a mutant copy of the *Arabidopsis* MYB transcription factor gene *MALE STERILITY 188* (*MS188*/*MYB80*) show precocious PCD [74]. MS188 regulates the A1 aspartic protease UNDEAD. This protease negatively regulates PCD, possibly by hydrolysing PCD-inducing protein(s) in mitochondria [74]. Mitochondria play an important role in PCD in animals and may also function in plant PCD. Cytochrome c release from mitochondria, a signature of animal PCD, is also detected in plants [75]. The exact

role of mitochondria in tapetal PCD is still unknown but mitochondria have been observed to persist throughout PCD [75].

The production of reactive oxygen species (ROS) also appears to play a role in the regulation of PCD [76]. In the rice *mads3* mutant, PCD occurs prematurely and ROS-homeostasis genes are mis-expressed leading to a loss of characteristic changes in ROS level [77]. The timing of ROS production appears to be an important factor for PCD and many ROS-producing *RESPIRATORY BURST OXIDASE HOMOLOGUE (RBOH)* genes are targets of tapetal transcription factors [78].

Altering the timing of PCD often leads to male sterility, which in crops has been used in breeding to control gene flow. Genes that control PCD in the tapetum can be important agronomic targets.

1.1.7 Tapetum transcriptional network

Genetic screens for male sterile mutants have led to the discovery of many genes functioning in the tapetum. This includes transcription factors which show severe mutant phenotypes due to the disruption of large numbers of downstream targets.

The best studied of tapetum transcription factors are those downstream of DYSFUNCTIONAL TAPETUM I (DYT1). DYT1 is a basic helix-loop-helix (bHLH) transcription factor that functions early in tapetum development, from anther stage four (Fig. 1-6) [79-81]. DYT1 can act as both a homo- and heterodimer with other bHLH transcription factors; bHLH010, bHLH089, bHLH091 [81, 82]. Mutant *dyt1* plants exhibit abnormal anther morphology from anther stage 4 and their tapetal cells have enlarged vacuoles, lacking the densely stained cytoplasm typical of tapetum. Meiocytes in *dyt* plants are able to complete meiosis I but degenerate before cytokinesis [79]. Therefore, DYT1 is an early regulator of tapetum function, controlling many transcription factors and tapetal metabolic pathways [81].

DYT1 directly regulates the MYB transcription factor DEFECTIVE in TAPETAL DEVELOPMENT and FUNCTION1 (TDF1, also called MYB35)[83, 84]. TDF1 is

proposed to regulate callose and fatty acid metabolisms, as well as pollen wall development [83-85]. TDF1 directly regulates, and interacts with, the bHLH transcription factor ABORTED MICROSPORES (AMS) [80, 84, 85]. The TDF1-AMS complex activates downstream genes in a feedforward loop to regulate transcription supporting anther growth and development [86]. AMS expression occurs in a biphasic pattern, which may be due to AMS competitively forming protein complexes with other transcription factors. [87]. Different AMS-containing complexes may therefore drive progression of the tapetum through different cell states. Later expression of *AMS* drives transcription of sporopollenin biosynthesis genes and the MYB transcription factor gene *MS188* [80, 88]

MS188 is predicted to upregulate callose-degrading enzymes, such as *A6*, as well as regulators of PCD and sporopollenin biosynthesis genes [74, 89, 90]. MS188 also forms a complex with AMS to drive expression of the sporopollenin-biosynthesis gene *CYP703A2* [91]. MS188 may act partially redundantly with late stage transcription factor MYB99 [92]. Other transcription factors, such as MYB32 and MYB4, are known to function at these stages to regulate sporopollenin biosynthesis, but their interactions with other factors is poorly understood [93].

MS188 is upstream of the PHD-finger transcription factor MALE STERILITY 1 (MS1) [94]. MS1 functions late in tapetal development to regulate sporopollenin biosynthesis and PCD [95, 96]. Degeneration of the microspores occurs soon after their release from the tetrad in *ms1* mutants, which also have vacuolated tapetal cells [94]. Interestingly, MS1 is expressed only in the tapetum for a short period between late tetrad stage and microspore release, suggesting that MS1-regulated functions at this stage are essential for pollen development [96, 97]

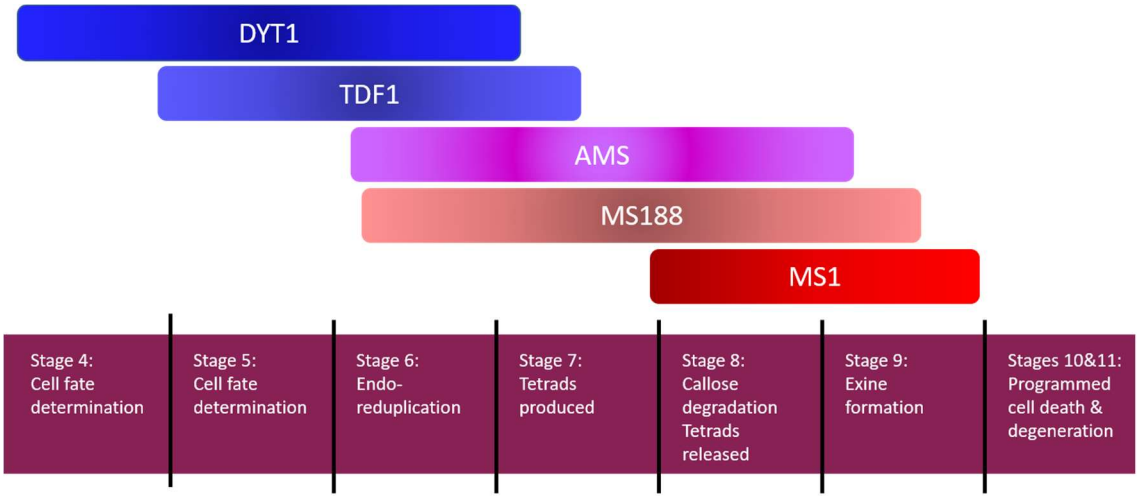


Figure 1-6: A representation of transcription factor expression through tapetal development. DYT1 is expressed early in tapetal development and upregulates TDF1. TDF1 regulates callose and fatty acid metabolism. TDF1 upregulates AMS and they form a complex to drive expression of downstream targets. AMS expression occurs in two phases. AMS regulates and forms a complex with MS188 at later anther stages to drive expression of sporopollenin-biosynthesis genes. MS188 also drives expression of the PCD-inhibitor UNDEAD. MS1 is downstream of MS188 and acts late in tapetal development to regulate expression of sporopollenin-biosynthesis and PCD genes. Expression data from [80].

Together, these transcription factors form the central network of tapetum development. Other transcription factors have been identified which regulate tapetum biology, but how they fit with this tapetal central network remains to be understood.

Summary

Male germline development relies on a wide range of function performed by the tapetum to produce mature pollen. The tapetum functions in the early development of germlines by supporting meiocytes. After meiosis, the tapetum controls the degradation of callose around tetrads and builds the exine walls around microspores. For pollen development to complete, the tapetum must perform programmed cell death, releasing pollen coat material onto the pollen grains. These processes are controlled by genes forming a core transcriptional network. As the tapetum develops the sexual lineage also undergoes large-scale changes in chromatin organisation and reprogramming of epigenetic marks, which are reviewed below.

1.2: Plant epigenetics and sexual lineages

Germline development is an important window for epigenetic reprogramming. Cells must transition from a somatic to a germline cell fate and the epigenetic marks laid down at this time can be heritable. Transposons must also be controlled, as new insertions will be inherited and so affect offspring fitness. Organisms are locked in an evolutionary arms race with transposons, so epigenetic mechanisms to control their expression in the germline are constantly evolving.

1.2.1 Transposons

Transposons are DNA sequences that can change their location in the genome and selfishly replicate at the expense of the host genome. The mutagenic capacity of transposons, and their ability to affect phenotype, has been thoroughly characterised since their discovery [98]. Transposons are ubiquitous in eukaryotic genomes and vary in their sizes, distributions and modes of replication. Two broad classifications of transposons exist based on the transposition method employed. Retrotransposons (Class I) replicate via an RNA and cDNA intermediate, while DNA transposons (Class II) can be excised and integrated into new locations by a transposase [99, 100]. Transposons exist as autonomous or non-autonomous elements. Autonomous transposons encode the enzymes required for their transposition, while non-autonomous elements possess *cis*-acting sequences which can be recognised by transposition factors supplied *in trans* [99].

Retrotransposons can be further classified by the presence or absence of long terminal repeats (LTRs). In *Arabidopsis*, the predominant LTR retrotransposons are *GYPSY* and *COPIA* elements and the predominant non-LTR retrotransposons are *LINE* elements [99]. *Arabidopsis* DNA transposons can be broadly classified into 6 classes; *MARINER, Tc1, POGO, HARBINGER, hAT, EnSpm/CACTA, MuDR*, and the rolling circle *HELITRON* elements [99].

Most transposon families are enriched in the pericentromeric regions of chromosomes, but *GYPSY* and *EnSpm/CACTA* elements show an enrichment for centromeres [99]. Many DNA elements, such as *HELITRON* and *MUTATOR* elements show high abundances in proximity to genes [101-103]. In the case of *HELITRON* elements this is partially driven by preferential insertion into TA dinucleotides [102]. The genomic location of transposons affects the methods employed to maintain their silencing.

While transposons are largely a mutagenic force in the genome, their movement and regulation can also be a creative force in evolution [100]. Transposition can sometimes move flanking sequences along with transposons, leading to the movement of genes and the shuffling of regulatory elements. This can give rise to novel regulatory interactions and expression patterns [100]. Transposons that move in response to certain environmental stimuli, *e.g.* high temperature, can also confer stimulus-responsive regulation to neighbouring genes [104]. Epigenetic regulation of transposons can therefore affect gene expression in response to specific environmental conditions.

In *Arabidopsis*, transposons are efficiently silenced through much of the life cycle. Certain developmental timepoints, however, show a relaxation of transposon silencing and the expression of a large number of elements [105]. In the male sexual lineage, this occurs in both the meiocytes [106, 107] and mature pollen [108]. While transposon expression can give rise to small RNAs that silence transposons [109], it remains unknown whether this is the ultimate reason for transposon expression or simply a response to transposon activity [105].

1.2.2 Histone modifications

Expression of genes and transposons can be regulated through changes in chromatin; the structure of DNA, RNA and proteins that form chromosomes. DNA is wrapped around nucleosomes, an octamer of the core histones H2A, H2B, H3 and H4 [110]. Covalent modifications made to these histones change chromatin structure, affecting

DNA accessibility, protein binding, and transcription [111]. Chromatin can be divided into two broad classes: euchromatin and heterochromatin. Heterochromatin is associated with transposons and repetitive sequences, is relatively condensed, and is enriched at centromeres and telomeres. Euchromatin is less condensed and more gene-rich. Heterochromatin forms a transcriptionally-repressive environment and is associated with repressive histone marks.

Methylation of histone 3 at lysine 9 (H3K9me) is a conserved histone modification across eukaryotes that acts as a repressive mark for transcription. H3K9 can be dimethylated (H3K9me2) in plants by the action of SU(var)3-9 homologues SUVH4/KRYPTONITE (KYP), SUVH5 and SUVH6 [112-114]. Trimethylation of histone H3 at lysine 27 (H3K27me3) is another repressive mark associated with genomic silencing that is deposited by polycomb repressive complex 2 (PRC2) [115]. In contrast to these marks H3K4 methylation acts as a transcriptionally permissive mark that is recruited to transcriptionally active genes [115]. As well as histone modifications, chromatin can be modified through the incorporation of histone variants. Variants of the core histones can be resistant to certain histone marks, and so affect the formation of chromatin domains [116]. The linker histone H1 binds outside of the nucleosome and is important for the establishment of heterochromatin [117]. H1 is associated with dense heterochromatin and impedes the access of proteins such as polymerases and DNA methyltransferases [118]. The chromatin remodeller DECREASE IN DNA METHYLATION 1 (DDM1) is required for access to H1-containing heterochromatin [118].

Histone marks are recognised by specific protein domains, and so can recruit proteins and other chromatin modifiers to specific genomic regions. Feedback loops with histone modifications leads to maintenance of clearly delineated chromatin domains [111, 119]. Histone marks associated with heterochromatin are also strongly linked to another repressive mark, DNA methylation.

1.2.3 DNA methylation

In plants, DNA methylation predominantly occurs at transposons and other repetitive sequences [120]. While animals predominantly possess DNA methylation at symmetric CG dinucleotides, DNA methylation occurs in all sequence contexts in plant genomes [121]. The pathways employed to establish and maintain DNA methylation differ between sequence contexts. Methylation at symmetric CG sites is maintained by METHYLTRANSFERASE 1 (MET1) [122]. At CG sites, cytosines on both strands are usually methylated. After DNA replication, each new DNA molecule inherits one methylated cytosine while the newly synthesised strand will be unmethylated. This hemimethylation is detected by VARIANT IN METHYLATION (VIM) proteins, which recruit MET1 to methylate the newly synthesised DNA [123]. By this method, CG methylation can be reliably maintained across cell divisions.

While MET1 maintains CG methylation across the whole genome, the maintenance of methylation at CHG and CHH (where H = A, T, or C) sites depends on genomic location and chromatin environment. Due to the presence of H1 in heterochromatin, the chromatin remodeller DDM1 is required for DNA methyltransferase genes to access and methylate DNA [111, 118]. Non-CG methylation at heterochromatic loci is maintained by a reinforcing loop linking DNA and histone methylation (reviewed in [124]). The chromo- and BAH domains of CHROMOMETHYLASE 2 (CMT2) and CMT3 recognise H3K9me2 marks and methylate cytosines [124]. CMT3 shows a preference for hemimethylated CHG sites, and as such is considered a maintenance methyltransferase [125]. CMT2 shows a preference for unmethylated DNA, and so acts as a *de novo* methyltransferase at both CHG and CHH sites [126]. Owing to the symmetrical methylation of CHG sites, following DNA replication, methylation can be propagated to the newly synthesised strand via CMT3. Asymmetric CHH sites have no methylation on the opposite strand, so methylation requires the constant action of CMT2 for maintenance. Non-CG DNA methylation recruits histone methyltransferases in a self-reinforcing loop to maintain both DNA methylation and

H3K9me2 [126]. Genome-wide removal of either H3K9me2 or non-CG DNA methylation leads to a loss of the other [126, 127].

RNA-directed DNA methylation

In the more open chromatin environment of transposon edges and euchromatin, DNA methylation is maintained by the RNA-directed DNA methylation pathway (RdDM) (reviewed in [128]). The RdDM pathway acts via small RNAs (sRNA) to methylate transposons and consists of two plant-specific Pol II homologues, Pol IV and Pol V [128]. The Pol IV pathway controls the production of sRNA from methylated RdDM-target loci (Fig. 1-7). Pol IV transcribes DNA, producing short nascent RNA transcripts [129]. Pol IV transcription can be directed to specific loci by chromatin re-modellers such as CLASSY 1-4 [130], and the homeodomain protein SAWADEE HOMEODOMAIN HOMOLOG 1 (SHH1)[131, 132]. SHH1 can bind to H3K9me2, and unmethylated H3K4 marks to recruit Pol IV and initiate the RdDM pathway [131]. Pol IV transcripts are acted on by RNA-DEPENDENT RNA POLYMERASE 2 (RDR2) to produce double stranded RNA (dsRNA) (Fig. 1-7) [133]. dsRNA is cleaved by DICER-LIKE 3 (DCL3), and methylated by HUA ENHANCER 1 (HEN1), to produce mature 24 nt sRNAs (Fig. 1-7) [134-136]. These Pol IV-derived small RNAs are then able to move within and between cells to their site of action in the Pol V pathway (Fig. 1-7) [128].

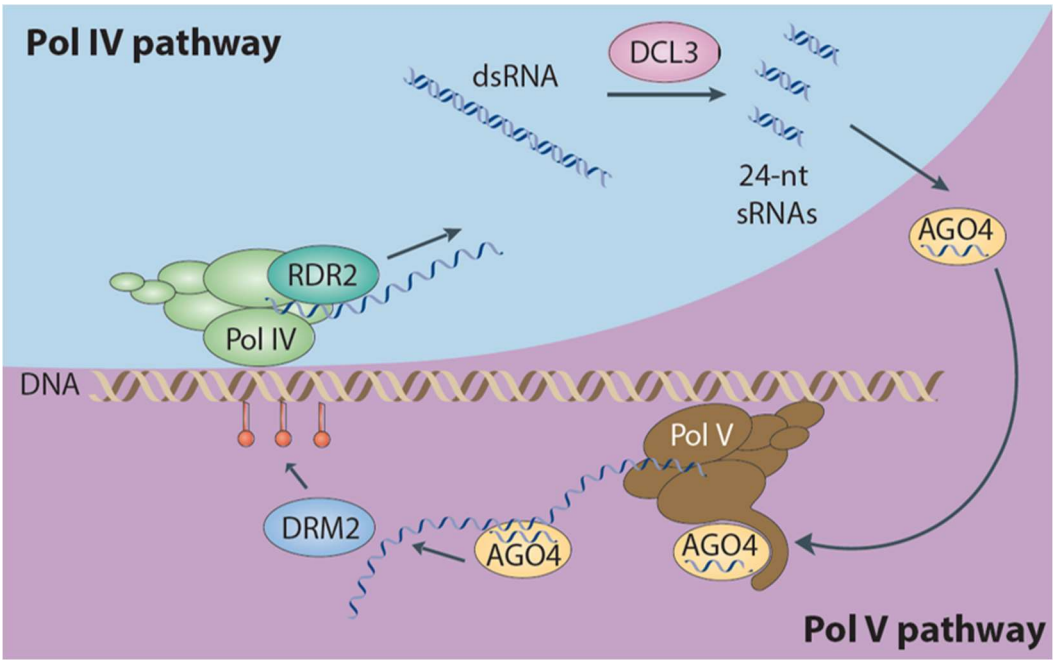


Figure 1-7: A simplified view of the RNA-directed DNA methylation pathway. RNA polymerase IV is recruited to DNA and transcribes short RNA transcripts. RNA is made double stranded by RNA-DEPENDENT RNA POLYMERASE 2 (RDR2) and diced into 24 nt sRNAs by DICER-LIKE 3 (DCL3). 24 nt sRNAs are then loaded onto ARGONAUTE proteins, most commonly AGO4. Pol V recruits AGO4 to sites of transcription, where the Pol V transcript acts as a template for AGO4 to direct DNA methylation by the DOMAINS REARRANGED METHYLTRANSFERASES (DRM1 and DMR2). Adapted, with permission, from [121]. Reproduced with permission from Hongbo Gao.

The Pol V pathway directs DNA methylation at sites that are complementary to the Pol IV-derived sRNAs. 24 nt sRNAs can be bound by ARGONAUTE proteins (most commonly AGO4) and recruited to their site of action by Pol V transcription (Fig. 1-7) [137]. Other members of the AGO4-clade also participate in RdDM, such as AGO6 in root and shoot apical meristems [138] and AGO9 which functions in reproductive cells [139]. Pol V has an extended carboxy-terminal domain containing GW/WG repeats, which forms an AGO-hook region to bind AGO proteins (Fig. 1-7) [140]. DNA methylation is then directed to sites complementary to 24 nt sRNAs and catalysed by the methyltransferases DOMAINS-REARRANGED METHYLTRANSFERASE 1 (DRM1) and DRM2 (Fig. 1-7). DRM2 is the predominant methyltransferase in somatic cells [127], though DRM1 is expressed in mature egg cells [141] and meiocytes (unpublished data). DRM methyltransferases methylate cytosines in all sequence contexts, though with a preference for CHH sites [142] Both Pol IV and Pol V are recruited to methylated DNA, so the RdDM pathway forms a self-reinforcing loop [131, 143, 144].

As well as the canonical Pol IV-RDR2 sRNA pathway, there are also other sRNA entry points to DNA methylation. RDR6 can act on Pol II transcripts to produce double stranded RNA which are processed into 21-22 nt sRNAs by DCL2/DCL4 [145-147]. RDR6-dependent sRNAs can act in post-transcriptional gene silencing but they are also able to direct transcriptional gene silencing via AGO6 [148]. These sRNAs can direct AGO6 to the chromatin of transcriptionally active transposons and establish DNA methylation. This pathway appears to be most active in reproductive tissue precursor cells to establish methylation at active long centromeric transposons [148]. RDR6 and AGO6 are also required to establish DNA methylation at new invasive transposons [149].

DNA hypomethylation caused by stress conditions, or during developmental reprogramming, leads to an increase in expression of a nuclear-targeted DCL4 variant. Nuclear localised DCL4 then acts with the Pol IV pathway to produce 21 nt sRNAs, from RdDM target loci, to act in post-transcriptional silencing [150]. This shows that alternative pools of small RNAs can not only drive DNA methylation but also be produced from methylated loci, particularly under stress. These examples highlight the abundant crosstalk between sRNA pathways and shows that plants have evolved a great deal of flexibility in tackling transposon activity.

An important aspect of the RdDM pathway is that the signal directing DNA methylation, the sRNA, is mobile. This allows sRNAs to act in *trans* within the same cell as well as between cells [151-156]. Changes in one cell type can thus affect the epigenetic state in another.

DNA demethylation

The mobile nature of sRNAs, and the self-reinforcing loops of DNA methylation, histone modifications, and sRNA production means that DNA methylation can spread to adjacent loci. To prevent the spread of DNA methylation, plants require mechanisms to remove methylated cytosines. This is performed by a family of 5-methylcytosine DNA glycosylases consisting of DEMETER (DME), DEMETER-LIKE

1/REPRESSOR OF SILENCING 1 (DML1/ROS1), DML2, and DML3. DME is the best studied of the family and largely functions during reproductive development [157-159]. The DME family remove 5-methylcytosine via a base excision repair mechanism. Thymine mismatched to Guanine in DNA can also be removed, as this can arise from the spontaneous deamination of methylated cytosines [160]. These DNA glycosylases therefore act to prevent the spreading of DNA methylation to open chromatin but also partially counter the mutagenic effect of cytosine methylation.

1.2.4 Epigenetics in the sexual lineage

In plants, germlines develop from somatic cells in the flowers after the switch to reproductive growth. Reproductive development is divided into three phases: sporogenesis, gametogenesis, and embryogenesis; the development of meiotic cells, gametes, and diploid progeny respectively. Different epigenetic reprogramming events are associated with each phase, and those affecting the male sexual lineage are detailed below.

Epigenetics in sporogenesis

Accompanying the switch from somatic to germline fate, *i.e.* during sporogenesis, are dynamic changes in chromatin [161, 162]. The development of spore mother cells in male and female organs (meiocytes and megaspore mother cells (MMCs), respectively) is associated with enlargement of nuclei and decondensation of chromatin [163, 164]. This decondensation is associated with a reduction in the linker histone H1 [163, 164]. In the MMC, there is widespread nucleosome remodelling with the incorporation of specific histone variants and a reported turnover of the centromere-specific H3 variant, CENH3 [164, 165]. In both meiocytes and MMCs, a transcriptionally permissive state is established, which corresponds with increases in the activating H3K4me3 histone mark and reductions of repressive H3K27me1, H3K27me3, and H3K9me1 marks [163, 164]. Chromatin decondensation in the male meiocytes leads to the wide-spread expression of transposons [106, 107, 163]. DNA methylation is also redistributed during sporogenesis as meiocytes show reduced

CHH methylation [9]. CHH methylation remains low throughout the sexual lineage [152, 154], but methylation at transposons is maintained through higher efficiency maintenance of methylation at CG sites by MET1 [166].

Small RNA pathways have been shown to function in sporogenesis in a range of plant species. The rice argonaute protein, MEIOSIS ARRESTED AT LEPTOTENE1 (MEL1), controls the switch from a mitotic to meiotic fate and homologous chromosome synapsis in early meiosis [167, 168]. MEL1 binds 21 nt phased small interfering RNAs (phasiRNAs) derived from somatic cells [167, 169]. In rice and maize, the tapetum is a source of 24 nt phasiRNAs, which are produced by the grass-specific DICER, DCL5 [156]. However, the exact function of reproductive 24 nt phasiRNAs remains largely unknown [170]. In *Arabidopsis*, somatic nurse cells regulate germline development through sRNA pathways. In the MMC, AGO9 and RDR6 control female gamete formation through signalling from the somatic nucellar cells [139]. Together, these data suggest that sRNA signalling from somatic cells is a conserved feature of germline development in plants [161, 171, 172].

Somatic production of sRNAs for the sexual lineage echoes the function of germline sRNA pathways in animals. The PIWI-interacting small RNA (piRNA) pathway has evolved to drive the formation of heterochromatin at transposons in the germline through the animal-specific PIWI clade of Argonaute proteins (reviewed in [173]). piRNAs are transcribed from piRNA clusters in the germline and surrounding nurse cells, through heterochromatin-specific polymerase machinery [174]. Nurse cell-derived piRNAs then act to silence transposons in the germline.

Epigenetics in Gametogenesis

After sporogenesis, meiocytes undergo meiosis and give rise to four haploid microspores. Microspores divide by mitosis to yield a large vegetative cell encompassing a smaller generative cell. The generative cell then gives rise two sperm cells after another round of mitosis [175]. The vegetative and sperm cells show large differences in chromatin states and nuclear organisation [161]. Sperm cells possess

highly condensed chromatin, marked with high levels of H3K9me2 [176], and specific histone variants [177]. The vegetative cell chromatin is largely decondensed, lacking centromere identity and possessing low levels of H3K9me2 [178], though CHH methylation is restored above the level seen in other sexual lineage cells [152, 154].

While CHH methylation is restored globally in the vegetative cell, active DNA demethylation occurs at transposons through the action of DME [152]. DME acts in the vegetative and central cells to ensure proper fertilisation and endosperm development. In the female central cell, DME is required to establish imprinting, which is essential for endosperm and seed development [155, 157, 179]. The imprinting of genes is brought about by DME targeting adjacent transposons. DME largely targets short AT-rich euchromatic transposons but can also be directed to heterochromatin in the central cell by the FACT complex [180]. The demethylation of transposons in the vegetative and central cells is proposed to give rise to sRNAs via transposon expression. These sRNAs then move to the gametes (sperm and egg cells) to reinforce silencing at transposon sequences [109, 152, 154, 179]. The vegetative cell is a short-lived nurse cell that risks its own genomic integrity by demethylating transposons in order to enhance silencing of transposons in the eternal germline cells [152, 154].

Sexual lineage specific methylation

In addition to reinforcement of DNA methylation at transposon sequences, the male sexual lineage possesses unique patterns of DNA hypermethylation. In a recent paper by Walker & Gao *et al.*, two types of sexual lineage methylated loci controlled by RdDM were identified [9]. Sexual lineage-hypermethylated loci (SLHs) are associated with transposons (particularly *HELITRONs*, unpublished data) and retain low levels of methylation in somatic tissues, particularly in the CG context (Fig. 1-8) [9]. Sexual lineage-specific methylated loci (SLMs) show methylation in all contexts which is established *de novo* in the sexual lineage, and are more strongly associated with genes than are canonical RdDM targets (Fig. 1-8) [9]. Methylation at SLM loci can also affect gene expression and splicing. In *drm1;drm2* plants, the meiosis gene *MULTIPOLAR*

SPINDLE 1 (*MPS1*) loses methylation in an intron, leading to incorrect splicing and the production of a truncated protein. This truncated protein disrupts meiosis, giving rise to an increase in triads, rather than tetrads [9]. Similarly, while piRNAs evolved to control transposons in the germline, they have also been co-opted to control gene expression in both the germline and somatic cells [181]. It appears that RdDM acts as an analogous pathway in plants.

While the sexual lineage undergoes widespread changes in global and local chromatin state, far less is known about the tapetal nurse cells. Ubiquitination of H2B (H2Bub) has been shown to regulate PCD-related genes in the tapetum of rice [182]. However, given that H2Bub marks actively expressed genes generally [111], regulation of PCD genes may be a clear phenotype for a more general effect [182]. Epigenetic changes, including increases in DNA methylation, have also been suggested to accompany PCD in *Arabidopsis* tapetum [183]. Gene regulation downstream of MS1 has been predicted to function through chromatin-level reorganisation of gene clusters [184]. The tapetum epigenome remains poorly understood as nucleotide-resolution maps of DNA methylation or histone modifications are lacking. Such data is essential if we are to gain a deeper understanding of epigenetic processes in the tapetum, and how they affect sexual lineage development.

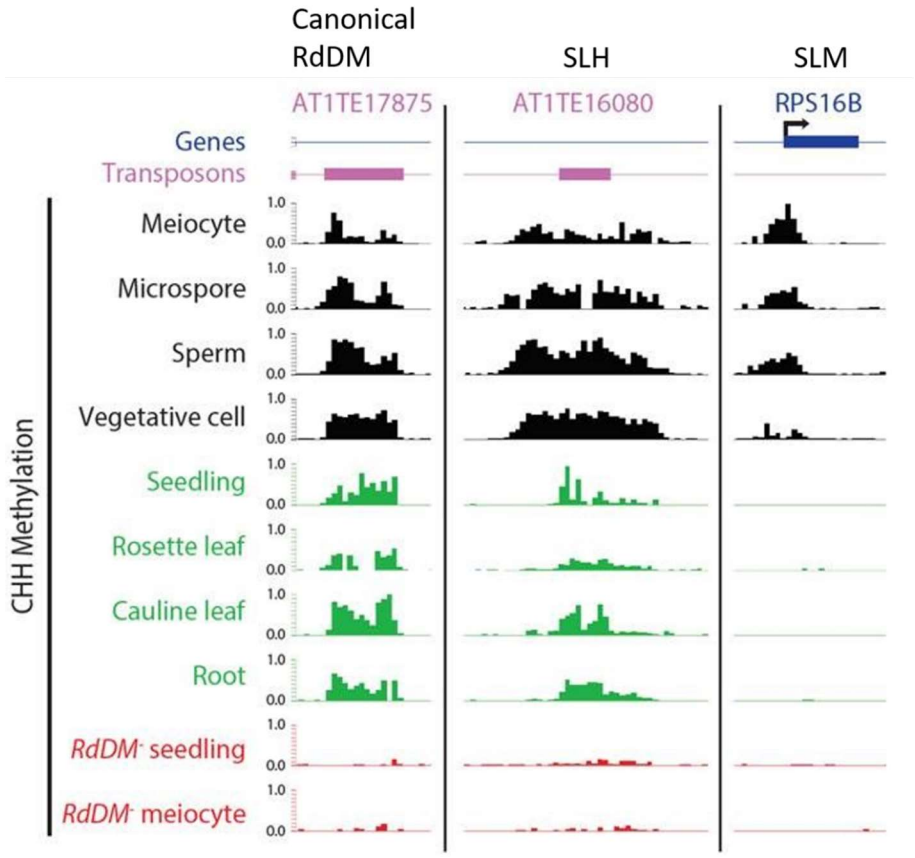


Figure 1-8: Methylation at canonical RdDM-targets, SLHs, and SLMs in the sexual lineage, somatic tissues and RdDM mutants. Canonical RdDM-targets are strongly associated with transposons and are methylated in all tissues. SLHs show hypermethylation in the sexual lineage but retain lower levels of methylation in somatic tissues. SLMs are methylated de novo in the sexual lineage and show no methylation in somatic tissues. SLMs are more strongly associated with genes than are SLHs and canonical RdDM-targets. Methylation at SLMs has also been shown to regulate gene expression [9]. Reproduced with permission from James Walker.

Summary

The precise control of DNA methylation and histone marks is required for the regulation of transposon and gene expression. As cells switch from a somatic to germline cell fate there is wide-spread reprogramming of DNA methylation and histone marks. This is accompanied by a wave of transposon expression. The control of transposons in germlines is under strong selective pressure, leading to the evolution of pathways to control them across a range of organisms. In many of these pathways, somatic nurse cells contribute sRNAs to aid the germline. While the tapetum is an essential cell type for male germline development, the epigenome of the tapetum is poorly understood.

1.3 Overview

In this thesis, I have taken a broad approach to understanding tapetum biology and its role in controlling male germline development. To achieve this, I have optimised a protocol to isolate the tapetum in *Arabidopsis* by fluorescence-activated cell sorting. Each chapter explores a different aspect of tapetum biology through the application of sequencing technologies to isolated tapetal cells. Access to whole genome sequencing data will allow researchers to build a holistic picture of tapetum function, putting expression of known tapetal genes into the context of co-expression, developmental timing, and chromatin environment. This data will facilitate further experimentation and modelling of the tapetum, as well as provide a more dynamic understanding of its role in male germline development.

In Chapter 2, I explore the tapetum transcriptome, with particular focus on the discovery of novel tapetal transcription factors. The tapetum is an important tissue of study for male fertility and many genes have been identified with tapetum functions. However, research has been hindered by the lack of tapetum transcriptomic data. The transcriptomes I produced provide an overview of tapetum transcription and an insight into tapetum-specific functions. This approach is extended in Chapter 3, where I use single-cell transcriptome sequencing to explore gene expression through the development of the tapetum. Developmental time is inferred from the transcriptomes of single-cells and has been used successfully to infer stage-specific expression for genes of interest. The tapetum performs many functions through discrete developmental stages. Therefore, a temporal understanding of gene expression is essential for the inference of gene function and regulatory relationships. The tapetum DNA methylome is explored in Chapter 4. Novel sites of DNA hypermethylation are characterised and similarities with sexual lineage methylation are explored. I test the hypothesis that the tapetum is a source of small RNAs for the sexual lineage and propose a model of tapetal sRNA transfer and action in the sexual lineage. Results from each chapter are discussed more generally in Chapter 5, with evidence from all chapters brought together to support hypotheses.

Chapter 2 Tapetum RNA-sequencing reveals novel genes regulating tapetal and germline development

2.1 Introduction

The control of germline development is important for both wild and domesticated plants. For the male germline, the tapetum is an essential cell layer which exerts a great deal of control over development. As such, the tapetum has been a target for research focussed on sexual reproduction for decades. Genetic screens have been able to identify large effect genes functioning in the tapetum, as male sterility is an easily identifiable phenotype. Transcription factors that regulate large numbers of genes in *Arabidopsis* tapetum have been identified and arranged into a transcriptional network [80]. Expanding this transcriptional network has been hampered by a lack of transcriptomic data from the tapetum specifically.

Within the anther locule the tapetal cells form a single cell layer surrounding the sexual lineage. As such, isolation of high-purity tapetum is extremely difficult. In rice [185] and *Arabidopsis* [85], laser capture microdissection has been employed. However, high purity cells are difficult to extract with manual techniques. Subtractive transcriptomics has been employed, using tapetum-deficient mutants to identify tapetum-enriched transcripts when compared to wild-type [82, 186]. However, given the crosstalk between the tapetum and sexual lineage, this method is prone to false positives.

In this chapter I have employed the fluorescence-activated cell sorting (FACS) approach to isolate *A. thaliana* tapetal cells for RNA-sequencing. I explored the tapetum transcriptomic data to discover novel transcription factors with putative tapetal functions. I then tested the functions of these transcription factors by

observing phenotypes of T-DNA insertional mutants. Using published data sets, genetic interactions were investigated and the tapetal gene regulatory network was expanded. Transposon expression in the tapetum was investigated and compared to somatic and sexual lineage tissues/cell types.

2.2 FACS isolation yields high purity tapetum

WT *A. thaliana* plants were transformed with a nuclear-targeted GFP fusion protein [187], under the control of a strong and tapetum-specific *A9* promoter [34]. Confocal microscopy of anthers showed strong GFP signals specifically in the tapetum throughout all stages of tapetal development (Fig. 2-1a). Protoplasts prepared from whole inflorescences were sorted based on GFP signal and forward scatter, using a Sony SH-800 cell sorter (Fig. 2-1b). Fluorescence microscopy of collected protoplasts showed GFP signal in 98% of the cells, indicating high purity (Fig. 2-1c). mRNA libraries were created using 400 protoplasts as previously described [188] and sequenced on a Hi-seq 4000 (Illumina) to a mean depth of 88 M reads. Biological replicates ranged from 88.5%-92.0% mapping efficiency (to the TAIR10 genome), suggesting all libraries were of good quality, despite being generated from just 400 cells each (Appendix Table A1).

I further assessed the purity of collected tapetal cells using the obtained transcriptomes. Known tapetal-specific genes such as *A9 [65]*, *MALE STERILITY 1* (*MS1*) [94],and *LESS ADHERENT POLLEN 5* [49], showed high expression in my RNA-seq data (Table 2-1). Consistent with the lack of chloroplasts in the tapetum, various photosynthetic genes, *e.g. CHLOROPHYLL A/B BINDING PROTEIN 3* (*CAB3*), were found absent or very lowly expressed (Table 2-1). The 1000 most highly-expressed genes show a significant enrichment for gene ontology (GO) terms associated with tapetal functions, such as “sporopollenin biosynthetic process” (GO:0080110), “acetyl-CoA biosynthetic process” (GO:0071616), and “pollen exine formation” (GO:0010584) [189-191]. These results demonstrate the high purity of collected tapetal cells.

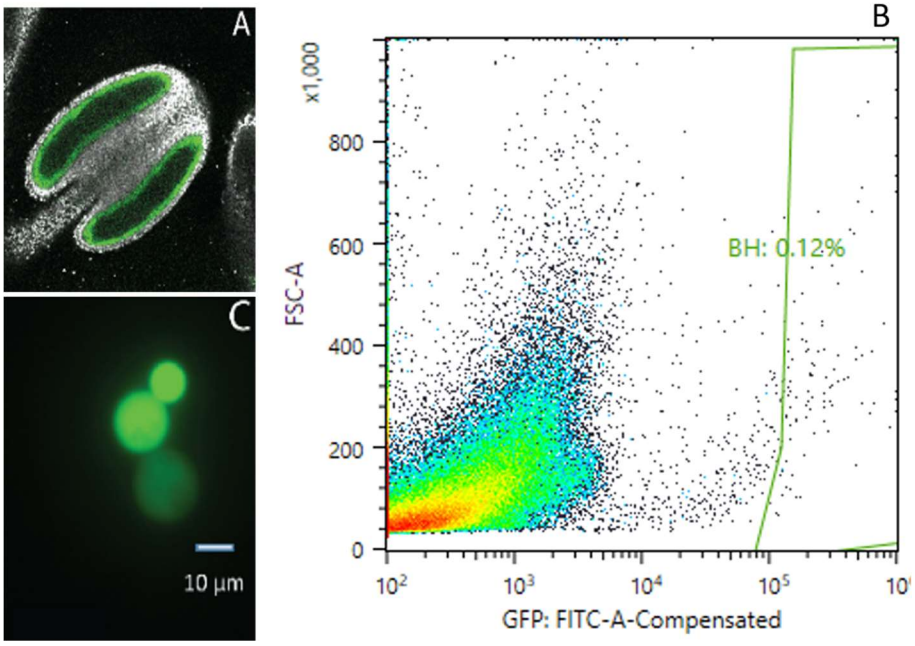


Figure 2-1: FACS protocol for the isolation of tapetal cells. Anthers show strong expression of a nuclear-targeted GFP specifically in the tapetum in *pA9::NTF* plants (**A**). After protoplasting, cells were sorted by their GFP-fluorescence above background levels (**B**). Fluorescence is roughly 1000-fold higher in tapetal protoplasts than non-tapetal protoplasts. In the flow cytometry plot the x-axis shows GFP fluorescence and the y-axis shows forward scatter, a measure of cell size. Sorted protoplasts (**C**) showed high purity, with >98% of cells sorted showing strong GFP-fluorescence.

A. thaliana tapetum transcriptomes have been published by Li *et al*. [85] using tapetum collected via laser microdissection from anthers of stages 6-7, and 8-10 [85]. Many genes (10201) were found differentially expressed between my transcriptomic data and combined transcriptomic data of early and late stage tapetum from Li *et al.,* [85] although most significant genes showed relatively small differences in expression (median difference = 16 Fragments per kilobase per million reads (FPKM)). In total, 1460 genes are significantly more highly expressed in my data at greater than 4-fold. My data showed an enrichment for genes known to function in the tapetum such as *LAP5* (Table 2-1). An upregulation of touch response genes, such as *TOUCH 2* (*TCH2*) and *TCH4* [192], is apparent in my data, presumably in response to mechanical disruption and protoplasting. A total of 3278 genes are significantly enriched in data from Li *et al., [85],* at greater than 4-fold. There is a significant enrichment for genes involved in photosynthesis (GO:0015979, p = 5.48x10^{-18}) suggesting contamination from non-tapetal tissues (Table 2-1). While data produced from FACS-sorted cells is of higher purity, there is also likely to be expression of genes in response to

protoplasting. Each method may therefore have advantages and disadvantages for specific analyses.

	Gene	FACS-sorted Tapetum	Li *et al.*, Tapetum Data	Leaf	Meiocyte
Tapetum-specific genes	A9	6517.19	2498.70	0.00	16.72
	TAP35	11958.90	1636.05	0.00	720.09
	A6	4682.26	872.58	0.17	37.95
	LAP5	3561.15	993.87	0.00	3.15
	MS1	29.80	7.51	0.00	0.09
Photosynthetic genes	LHCB6	3.18	4409.23	1679.50	5.90
	RBCS1A	10.87	8861.38	13769.70	9.36
	RBCS1B	0.20	600.81	1443.04	1.01
	CAB3	0.00	1462.98	2659.09	0.75
Meiotic genes	ASY1	1.74	11.29	0.23	188.94
	DMC1	4.42	52.30	1.47	247.68
	SDS	0.31	5.52	0.41	171.79
	ATSPO11-1	7.23	0.08	0.00	53.78

Table 2-1: Expression of known tapetum-specific, photosynthetic, and meiotic genes in my tapetum data, data from Li *et al.,* [85], rosette leaf and meiocytes (unpublished data). My tapetum RNA-seq data shows high expression of tapetum-specific genes, and an absence of expression of photosynthetic or meiotic genes. Data from Li *et al.,* shows high levels of photosynthetic gene expression, suggesting severe contamination of non-tapetal tissue. Expression is in FPKM.

2.3 Identification of tapetum-enriched transcription factors

To identify novel genes involved in tapetum development, I compared my tapetum RNA-seq data to that from whole anthers [82]. As a result, 917 genes were found significantly enriched in the tapetum relative to anthers ($p < 0.05$, Cuffdiff, fold change > 4). To narrow down this gene list, putative transcription factors were selected owing to their ability to affect the expression of large numbers of downstream genes. Genes associated with GO terms for "DNA binding" (GO:0003677), and/or "regulation of transcription, DNA-templated" (GO:0006355) were selected. I then further narrowed down the list to 11 genes by selecting those that are well expressed in the tapetum (FPKM > 15) and had not been associated with any tapetal function. T-DNA insertion mutants of these genes were analysed for tapetal and pollen phenotypes (Table 2-2).

Two *BASIC LEUCINE-ZIPPER* (*bZIP*) genes were found to be novel tapetum-enriched transcription factors. bZIP28 is an ER-bound transcription factor that regulates root elongation [193], and also functions in ER and heat stress [194]. bZIP28 is possibly upstream of THERMOSENSITIVE MALE STERILE 1 [195], functioning in the unfolded protein response [196]. These suggest that bZIP28 functions in the tapetum, but possibly only at times of stress. bZIP18 is predicted to be a homodimer [197], but downstream targets and the functions regulated remain unknown.

Gene ID	Description	SALK-line
AT1G02040	C2H2-type zinc finger family protein	SALK_021494
AT1G14350	MYB124, also known as FOUR LIPS (FLP)	SALK_033970C
AT1G17950	MYB52	SALK_138624C*
AT1G20670	DNA-binding bromodomain-containing protein	SALK_049806C*
AT1G28470	NAC010 also known as SECONDARY CELL WALL NAC DOMAIN PROTIEN 3 (SND3)	SALK_000287C*
AT1G76880	DF1	SALK_106258C*
AT1G18960	Myb-like HTH transcriptional regulator family protein	SALK_128611C*
AT2G40620	bZIP18	SALK_110712C*
AT3G10800	bZIP28	SALK_132285C
AT5G01380	GT3a	SALK_043542*
AT5G54310	NEVERSHED (NEV), also known as ARF-GAP DOMAIN 5 (AGD5),	SALK_118697*

Table 2-2: Selected transcription factors found to be enriched in the tapetum relative to whole anther tissue. SALK lines listed are those used for each gene throughout the chapter.

Two MYB transcription factors are also tapetum enriched. MYB124 functions in regulating the cell cycle in the stomatal lineage [198], but a tapetal function has not yet been described. MYB52 has been identified as a negative regulator of pectin demethylesterification in seeds [199] and of lignin biosynthesis in stems [200]. Pectin demethylesterification is an important process for the release of microspores from tetrads [44], and lignin biosynthesis shares many similarities with sporopollenin biosynthesis [42]. Therefore, MYB52 may act as a negative regulator of these processes in the tapetum.

The trihelix/GT-family DNA-binding factors are a relatively small and poorly understood family of transcription factors in *Arabidopsis*. Two members are tapetum-

enriched relative to whole anthers. *GT-3a* is a poorly studied transcription factor gene expressed in floral structures and the embryo sac [201]. The exact functions of GT-3a are unknown. Another trihelix transcription factor, DF1 regulates seed mucilage and root hair elongation [202, 203].The root hair phenotype of *df1-1* (SALK_106258) has been complemented through reintroduction of the native gene but reproductive phenotypes have not been reported [202].

Of the remaining genes, *AT1G02040* encodes a C2H2 Zn-finger protein predicted to be downstream of MALE STERILITY 188 (MS188) [204], but its exact function is unknown. *NAC DOMAIN CONTAINING PROTEIN 10/SECONDARY WALL-ASSOCIATED NAC DOMAIN PROTEIN 3 (NAC010/SND3*) functions in secondary cell wall biosynthesis and vascular differentiation [205, 206]. *NEVERSHED* (*NEV*) encodes an ARF-GAP DOMAIN protein that functions in leaf and floral abscission, and the *nev* mutant blocks organ shedding due to defects in membrane trafficking [207-209]. This may suggest a role in regulating tapetal membrane trafficking and excretion of pollen wall materials. No known function of AT1G18960 has been reported, and transcripts have only been detected in floral and anther tissues [210].

2.4 Identification of genes specifically expressed in the tapetum

I was also interested in finding genes that are expressed specifically in the tapetum, as these genes very likely encode for tapetum functions. To find tapetum-specific genes, comparisons were performed between FACS-isolated tapetum transcriptomes and eleven RNA-seq datasets from different *Arabidopsis* tissues and cell types. Genes were selected if they had a tapetum specificity score greater than 0.9 from CummeRbund [211] (Table 2-3).

This analysis identified 25 highly tapetum-specific genes (Table 2-3). These 25 genes include those known to function in the tapetum such as; *CYP703A2* [51], *UNDEAD* [74], *A7* [64], and *MS1* [94]. Several other genes were identified, such as the protease AT4G30030 which shows sequence similarity to UNDEAD [212] and may also

function as a negative regulator of tapetal programmed cell death [74]. To narrow down the list of tapetum-specific genes to investigate, only putative transcription factors were selected to match those selected as tapetum-enriched.

Several transcription factor genes were found to be tapetum-specific: *LOB DOMAIN CONTAINING 2* (*LBD2*), *bHLHc18*, and *NAC025*. Of these, *bHLHc18* and *NAC025* are only expressed at low levels (FPKM = 1 and 15, respectively). A T-DNA insertion mutant was investigated for *NAC025* (SALK_060459), but the plants contained multiple T-DNA inserts, and showed a fertility defect which did not segregate with the *nac025* insert. As such, only *LBD2* was investigated further as it showed high tapetum specificity, strong expression (FPKM = 358), and a mutant line is available with a single T-DNA insertion in the 3′ UTR (SALK_112456). It has also been suggested to be a downstream target of the tapetum-expressed transcription factor MS188 [204].

Gene Name	Description	Name	Expression level
AT1G01280	Cytochrome P450 family 703 subfamily A polypeptide 2	CYP703A2	359.57
AT1G06280	LOB domain-containing protein 2	LBD2	358.11
AT1G08065	Alpha carbonic anhydrase 5	ACA5	53.08
AT1G12540	bHLH DNA-binding superfamily protein	bHLHc18	1.02
AT1G23670	Domain of unknown function (DUF220)		3.86
AT1G28375			1107.98
AT1G33820			0.79
AT1G44224	ECA1 gametogenesis related family protein		11.59
AT1G50470	F-box associated ubiquitination effector family protein		2.196
AT1G61110	NAC domain containing protein 25	NAC025	14.89
AT1G68875	Hypothetical protein		44.68
AT1G75940	Glycosyl hydrolase superfamily protein	ATA27/ BGLU20	12.06
AT2G28725	Forkhead box protein G1		6.32
AT2G31215	bHLH DNA-binding superfamily protein		13.69
AT2G41415	Maternally expressed gene (MEG) family protein		192.37
AT3G06090	Peptide PIPL2	PIPL2	2.20
AT3G23770	O-Glycosyl hydrolases family 17 protein		395.76
AT3G26140	Cellulase (glycosyl hydrolase family 5) protein		2.361
AT3G55570	Cytoplasmic tRNA 2-thiolation protein		2943.60
AT4G12920	Eukaryotic aspartyl protease family protein	UND; UNDEAD	93.56
AT4G22870	2-oxoglutarate (2OG) and Fe(II)-dependent oxygenase superfamily protein; leucoanthocyanidin dioxygenase		6.625
AT4G28395	Bifunctional inhibitor/lipid-transfer protein/seed storage 2S albumin superfamily protein	A7	168.01
AT4G28397	Non-specific lipid-transfer-like protein		25.14
AT4G30030	Eukaryotic aspartyl protease family protein	(UND-like)	9.30
AT5G22260	RING/FYVE/PHD zinc finger superfamily protein	MS1	25.66

Table 2-3: Genes found to be tapetum-specific (specificity score >0.9) when compared to seedling, embryo, rosette leaf, meiocyte, pollen, and root epidermis, cortex, endodermis, stele, and columella (methods 2.11.5)

2.5 Tapetal defects were observed in mutants of *bZIP18*, *DF1* and *LBD2*

To discover the function of tapetum-enriched transcription factors, anther phenotypes of T-DNA insertion mutants were examined via cross-sectioning of resin-embedded inflorescences. The mutant lines examined are marked with asterisks in Table 2-2 and include the mutant of *LBD2* (SALK_112456), which was identified as tapetum-specific. Severe anther phenotypes were observed in three mutant lines - SALK_110712; *bzip18*, SALK_106258; *df1*, and SALK_112456; *lbd2* (Fig. 2-2).

In WT anthers, distinct stages of tapetal development can be seen by resin sectioning (Fig. 2-2). At stage 5 the tapetum forms a monolayer surrounding the meiocytes. At stage 6 the meiocytes are surrounded by a callose cell wall and progress through meiosis, yielding tetrad meiotic products at stage 7 (Fig. 2-2). The callose cell wall surrounding the tetrads is broken down and individual microspores are released by stage 8 (Fig. 2-2). A sporopollenin cell wall is rapidly produced by the tapetum and begins to coat the microspores by stage 9 [42] (Fig. 2-2). From stage 10 the tapetum becomes thinner and visible degeneration can be seen [18]. Tapetum degeneration is complete by stage 11 (Fig. 2-2).

The *bzip18* mutant (SALK_110712) shows the subtlest phenotype among the three mutants. At early stages (5-6) the tapetum appears normal in structure and staining (Fig. 2-2). At stage 7, Callose breakdown is unaffected, and microspores are successfully released from tetrads (Fig. 2-2). From stage 8 multi-layered tapetum can be seen in anthers (Fig. 2-2). At stages 8-9 misshapen tapetal cells, which expand into the locule slightly, were observed (Fig. 2-2). The timing of tapetum degeneration is unaltered, occurring at stage 11 (Fig. 2-2).

Anthers of the *df1* mutant show a similar, but more severe, phenotype to *bzip18*. From stage 8 multi-layered tapetum is observed, though this does not affect tetrad release (Fig. 2-2). At stages 8-9 tapetal cells appear disordered and do not form a contiguous cell layer in the anther (Fig. 2-2). Some tapetal cells are detached from the locule wall

and expand inwards to fill the locular space (Fig. 2-2). PCD of the tapetum at stage 11 progressed normally in the *df1* mutant (Fig. 2-2). Another T-DNA insertion mutant of a trihelix transcription factor, SALK_043542 (*gt-3a*), also showed tapetal defects to a similar level as *bzip18* (Appendix Fig. A1). However, as *DF1* is expressed to a higher level, and *df1* shows a more severe phenotype, only *df1* was investigated further.

In anthers of *lbd2*, early stages of anther development appear WT-like (Fig. 2-2). At stage 8 irregular tapetal cells are apparent and these may expand into the locules. From stage 9 the tapetum appears vacuolated and, especially where multi-layered, invades the locular space (Fig. 2-2). At early stage 9 in 6.5% of locules, the expanded tapetal cells fill up the entire locular space, leading to collapsed and misshapen microspores (Fig. 2-2). At late stage 9, tapetal cells seem to undergo precocious degeneration in 11.2% of the locules (Fig. 2-2). The majority of locules, however, can progress through stage 9 and show signs of tapetum degeneration normally at stage 11 (Fig. 2-2).

Genotyping of *df1*, *bzip18* and *lbd2* mutants detected insertions in exons or UTRs of corresponding genes (Fig. 2-3). The *df1-1* (SALK_106258) allele has previously been confirmed as a knockout [202]. Quantitative reverse transcription PCR (qRT-PCR) detected expression in *df1* (28% of WT level), though this is likely to result from a natural antisense transcript which overlaps *DF1* (AT1G76878). As the T-DNA insertion is at the start of exon 2, functional DF1 protein is unlikely to be produced in *df1*. qRT-PCR showed reduced transcription of *LBD2* and *bZIP18* in inflorescences of corresponding mutants, both at 22% of WT levels (Fig. 2-3). These results suggest the tapetum defects observed in these mutants are likely caused by the knock-down or knock–out effect of the corresponding gene.

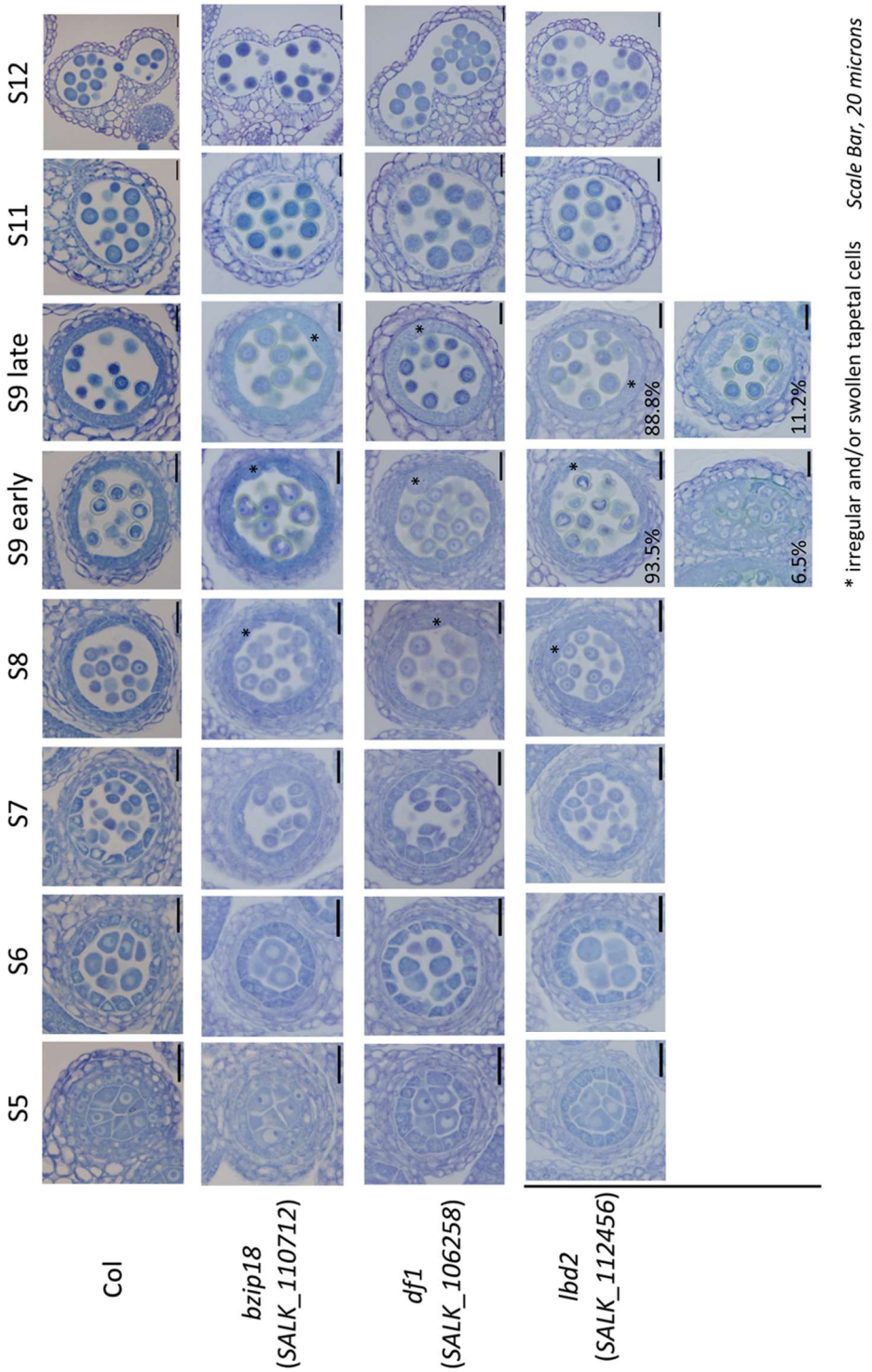


Figure 2-2: Cross sections of anther locules from WT, and *bzip18*, *df1* and *lbd2* mutants. All mutants show WT development from stage 5 to 7. At stage 8 irregular or swollen tapetal cells are observed in all mutants. This phenotype is exacerbated in mutants at stage 9. *lbd2* anthers show more severe tapetal phenotypes at stage 9. 6.5% of locules show tapetal cells expanded into the locule at early stage 9. By late stage 9 11.2% of locules show expanded tapetum which appears to degenerate early.

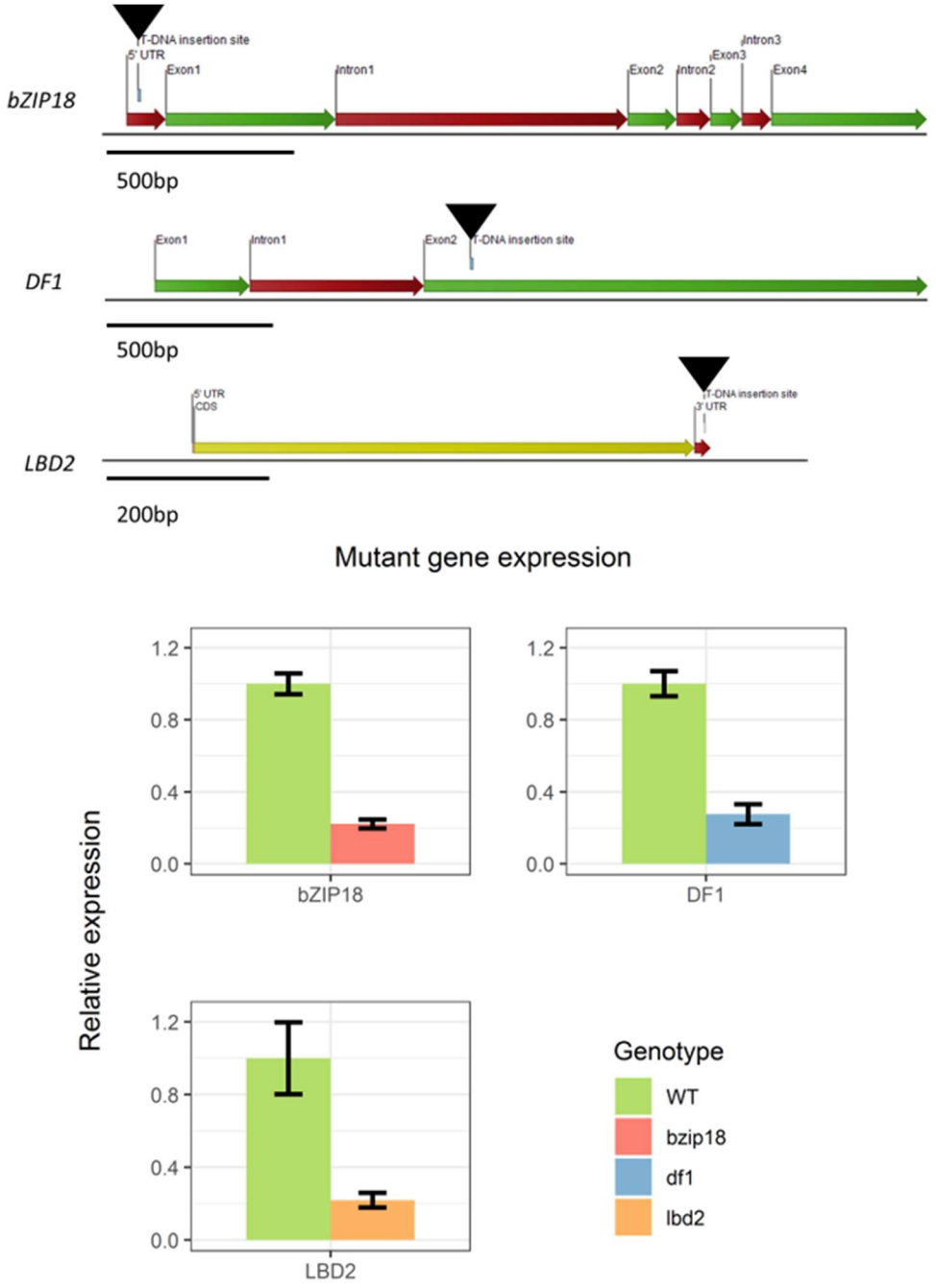


Figure 2-3: Diagrams highlighting the T-DNA insertion sites (black triangles), exons (green), introns and UTRs (red) of *bzip18* (SALK_110712), *df1* (SALK_106258), and *lbd2* (SALK_112456). LBD2 lacks introns and is represented by a single coding sequence. *bzip18* possesses a T-DNA insert in the 5' UTR, leading to a 78% reduction in expression. *df1-1* has a T-DNA insert at the start of exon two, leading to a gene knockout in roots [202]. Expression is detected in df1 inflorescences (28% WT), though this may be from a natural antisense transcript which overlaps DF1 (AT1G76878). *lbd2* has a T-DNA insert in the 3' UTR, leading to a 78% reduction in expression. Error bars represent standard error.

2.6 *df1* and *lbd2* mutant plants show pollen defects

While the tapetum in *bzip18*, *df1*, and *lbd2* shows varying degrees of defect, whether these translate into a reduction in fertility is not certain. As a measure of fertility, silique lengths of *bzip18*, *df1*, and *lbd2* T-DNA insertion mutants were measured (Fig. 2-4). Siliques of WT plants have a median length of 14.58 mm and a mean of 13.561 mm. Even in WT there is a negative skew in lengths, with several siliques not fully expanded (Fig. 2-4). All mutant lines show a lower median length that WT plants (*bzip18* 14.38 mm, *df1* 12.34 mm, *lbd2* 13.24 mm) but *bzip18* did not differ significantly from WT (p = 0.132, Conover-Iman test) (Fig. 2-4). Both *df1* and *lbd2* showed significantly shorter siliques than WT ($p < 0.001$ & $p < 0.01$ respectively) (Fig. 2-4). Siliques of *lbd2* did not differ significantly from *bzip18* or *df1* ($p > 0.025$) but siliques of *df1* were significantly shorter than *bzip18* ($p = 0.0001$) (Fig. 2-4).

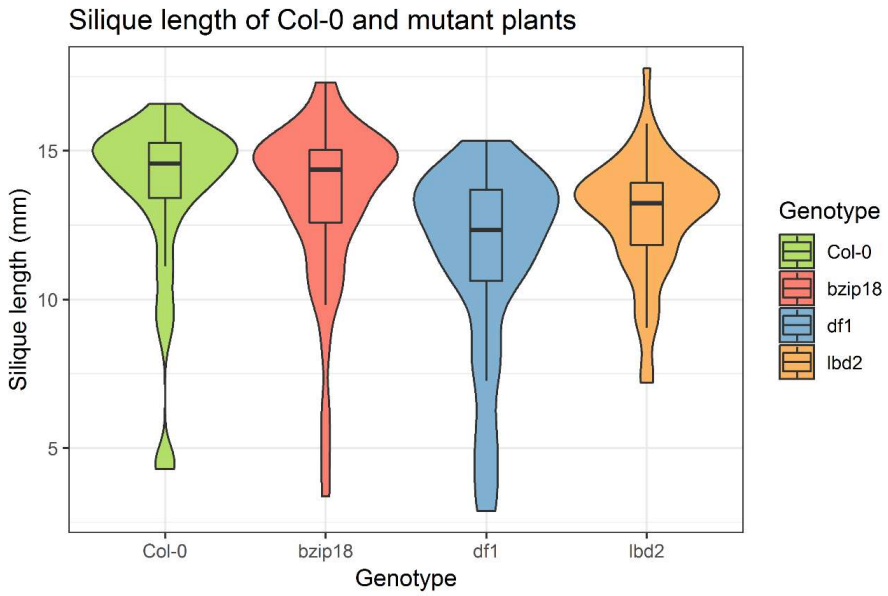


Figure 2-4: Silique lengths of WT and, *bzip18*, *df1*, and *lbd2* mutant plants. Distributions represent the lengths of 50 siliques, ten each from five plants. Siliques of bzip18 do not differ significantly in length from WT (p = 0.132, Conover-Iman test) but both df1 ($p < 0.001$) and lbd2 (p = 0.0016) have significantly shorter siliques than WT.

To investigate pollen viability in *bzip18*, *df1* and *lbd2* mutants, Alexander staining was performed [213] on WT and mutant anthers (Fig. 2-5). Viable and non-viable pollen are stained in red and blue, respectively, by Alexander staining. Similar to WT, pollen

of *bzip18*, *df1* and *lbd2* mutants were mostly stained red, suggesting single mutations of these genes do not affect pollen viability (Fig. 2-5).

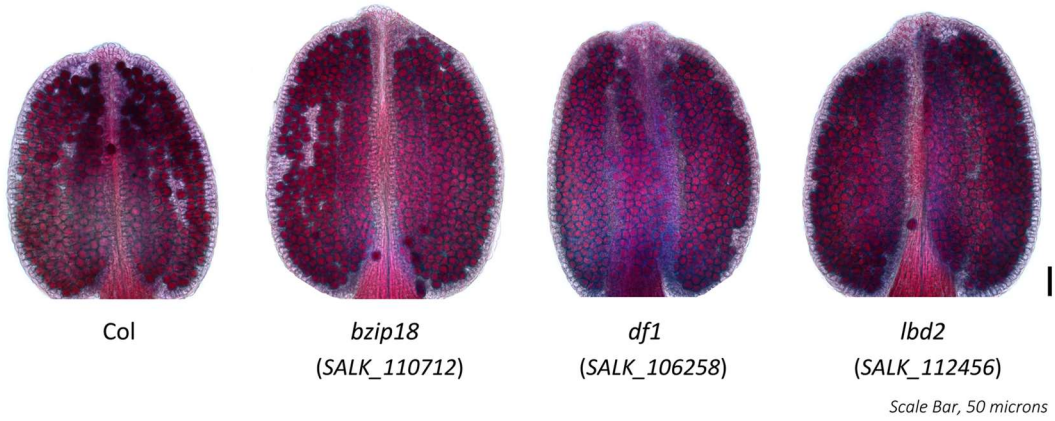


Figure 2-5: Alexander staining of WT, *lbd2*, *df1*, and *bzip18* plants. Viable pollen stain red while aborted pollen stain blue/green, due to differences in cytoplasm pH. No mutant plants showed an increase in aborted pollen relative to WT and almost all pollen stained red. Five anthers were stained for each genotype and those shown are representative. Scale bar represents 50 μm

Alexander staining assesses pollen viability by reacting to pollen cytoplasm pH, and cannot identify subtle defects in pollen morphology [213]. I further investigated pollen morphology in *WT, bzip18, df1 and lbd2* anthers via scanning electron microscopy (SEM). While some misshapen pollen could be seen from all WT and mutant anthers, the proportion of misshapen pollen was significantly higher in *df1* and *lbd2* mutants than WT (mean = 0.167, p = 0.0129 and mean = 0.194, p = 0.0109, respectively, comparing to mean = 0.090 in WT; Conover-Iman test) (Figs. 2-6 & 2-7). The proportion of misshapen pollen from *bzip18* was not significantly different from any other genotype ($p > 0.025$, Conover-Iman test). It is noteworthy that the difference from WT in *df1* and *lbd2* are largely driven by rare anthers with high proportions of misshapen pollen. For *df1*, three anthers show greater than 25% misshapen pollen. This is more severe in *lbd2* where five anthers show greater than 25% misshapen pollen and one anther showed >99% misshapen pollen (n = 168). This rare severe phenotype is consistent with the strong tapetal defect observed in a small proportion of locules at stage 9 (Fig. 2-2) and suggests that the tapetal defect may be the cause of pollen abnormality.

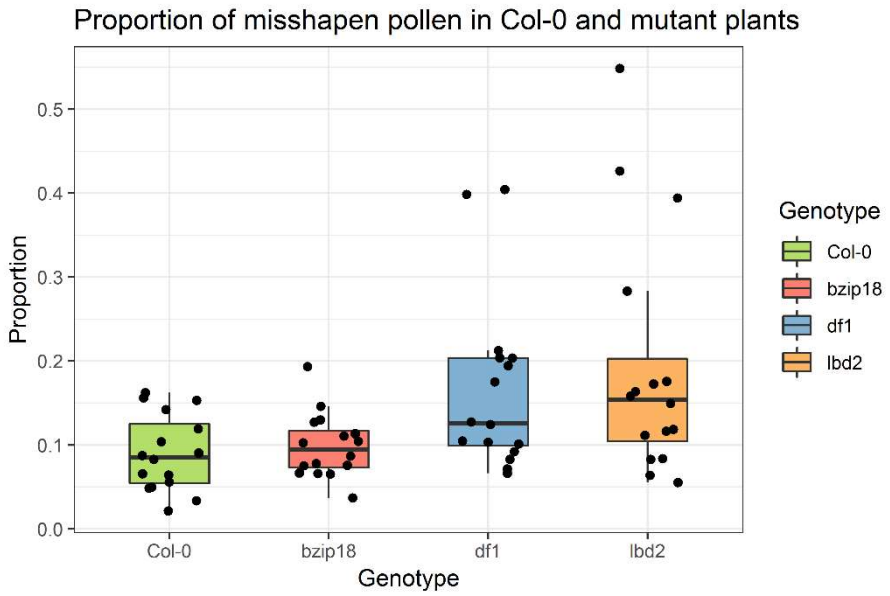


Figure 2-6: Rates of misshapen pollen observed in WT, *bzip18*, *df1*, and, *lbd2*. Malformed pollen was seen in all genotypes but in WT and *bzip18* this was consistently low. Rates of defective pollen did not differ between WT and *bzip18* but both *df1* and *lbd2* had significantly higher rates of defective pollen (p = 0.0129 and p = 0.0109, respectively). Data are rates of defective pollen from individual anthers.

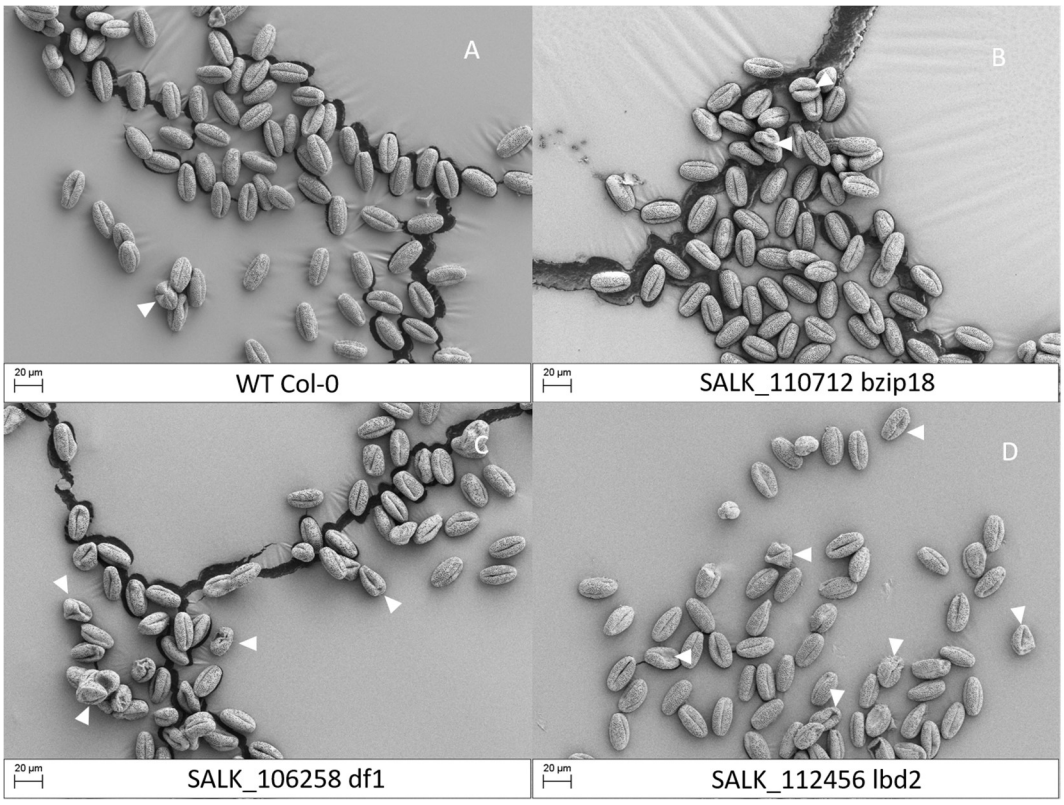


Figure 2-7: SEM images of pollen from WT (**A**), *bzip18* (**B**), *df1* (**C**), *lbd2* (**D**). The majority of pollen in all genotypes have no visible morphological defects. Reticulate exine pattern can be seen with a straight aperture running the length of the pollen grains. Example defective pollen grains are highlighted by white arrows. Pollen were classified as misshapen if their shape differed significantly from WT, the pollen grain appeared collapsed, or if the aperture was bent. Even in WT misshapen pollen are seen, but this occurs at a low rate. Scale bar 20 µm.

Misshapen pollen are seen in all genotypes, including WT. Misshapen pollen mostly show subtle defects in shape while not differing greatly in size or exine pattern from WT (Fig. 2-8b). In both *lbd2* and *df1*, anthers with high rates of misshapen pollen also show pollen with more severe phenotypes in pollen exine structure. In these pollen, pillars of sporopollenin (baculae) appear to lack caps (tecta), which form the stereotypical reticulate exine shape of WT. This would suggest that sporopollenin production is disrupted in these mutants.

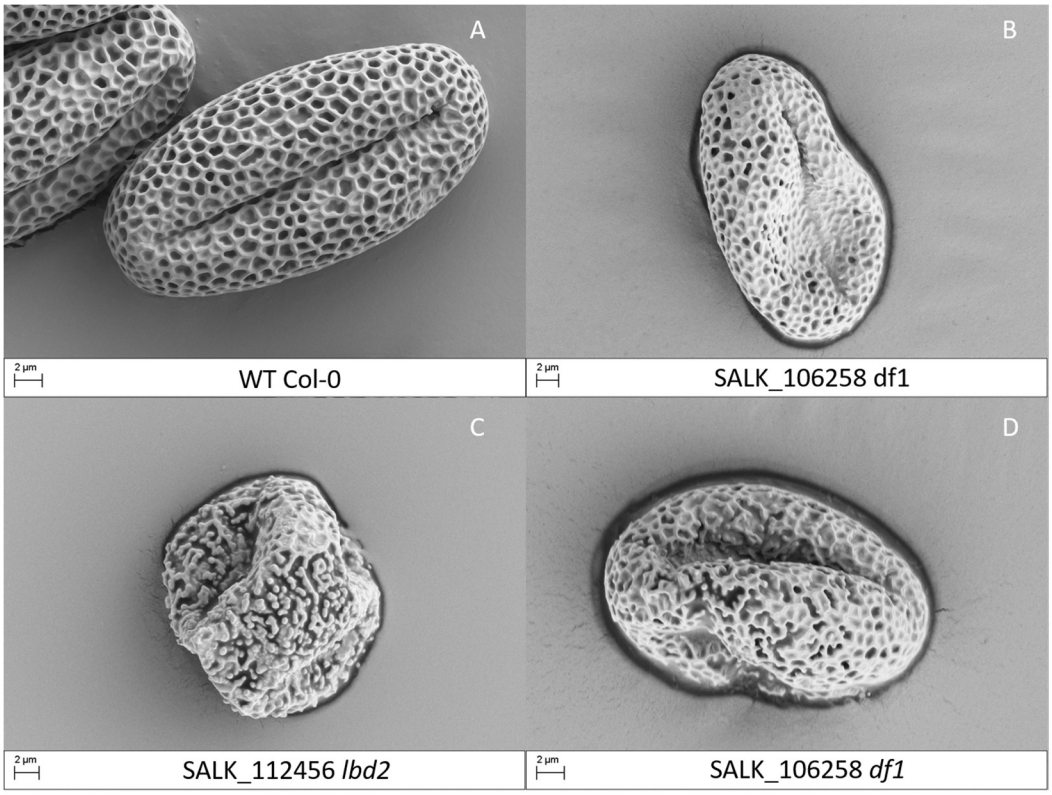


Figure 2-8: Magnified (12K x) SEM images of pollen from WT (**A**), *df1* (**B, D**), and *lbd2* plants. WT pollen show a stereotypical hexagonal exine structure on the pollen surface. In most misshapen pollen of *df1* (**B**), and other mutants, exine structure remains intact. Rarely, in pollen of *lbd2* (**C**) and *df1* (**D**), defects in exine structure are seen. Pollen appear to lack the sporopollenin caps (tecta) connecting pillars (baculae) into the exine structure seen in WT [41].Scale bar 2 µm.

DF1 and LBD2 belong to large transcription factor families that display redundancies among members [214, 215]. To explain the incomplete penetration of the tapetum and pollen phenotypes (Figs. 2-2 & 2-6), the expression of possible homologues of *DF1* and *LBD2* were explored (Fig. 2-9). DF1 is a trihelix/GT-family transcription factor which is most similar to GT2 and has been shown to function redundantly with GT2-LIKE 1 (GTL1) in root hair development [202], though *GTL1* is undetected in the

tapetum. The *GT-1* gene [216], of a different clade, is expressed to a similar level as *DF1* in the tapetum, suggesting that they may act redundantly. However *GT-1* appears to be ubiquitously expressed across most organs [210] (Fig. 2-9). The tapetum-enriched gene *GT-3a* (section 2.3) is also expressed, though not as highly as *DF1* (Fig. 2-9). Given the tapetal phenotype seen in the *gt-3a* mutant (Appendix Fig. A1), and the phenotypic similarity to *df1*, GT-3a may act partially redundantly with DF1 to regulate tapetal function.

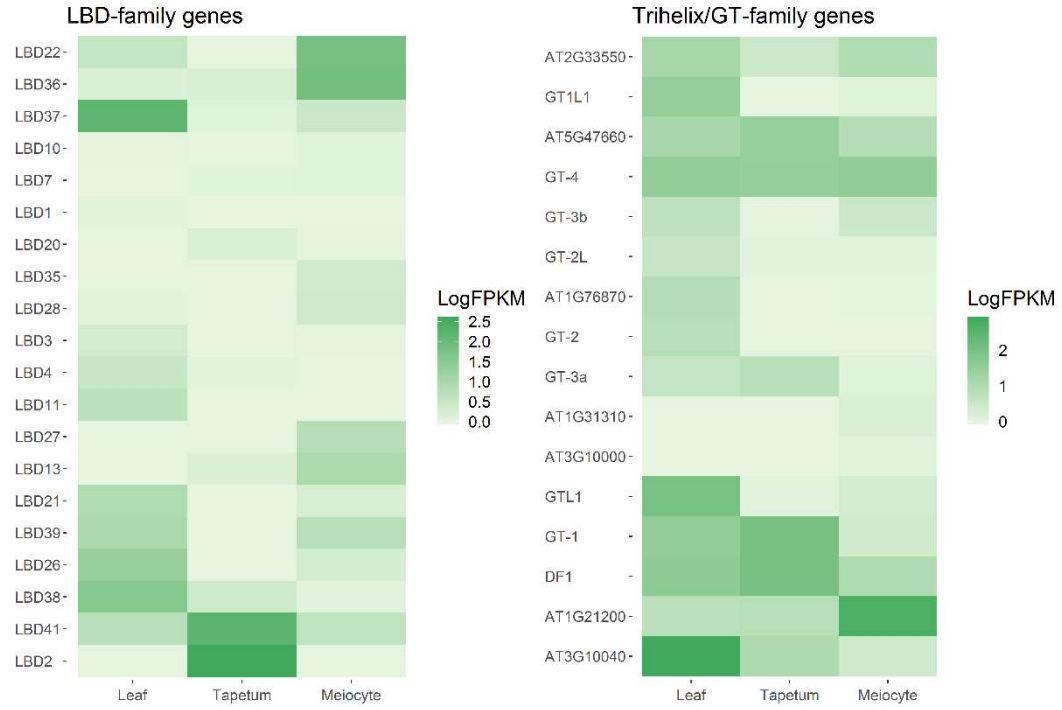


Figure 2-9: Expression of LOB-domain containing and trihelix/GT-family transcription factors in rosette leaf, tapetum, and meiocytes. *LBD2* is the most highly expressed LBD-family gene detected in the tissues tested. *LBD41* is the only other LBD-family genes expressed to a high level in the tapetum. LBD family members not expressed in any of the tissues selected were not included in the expression matrix. All trihelix family transcription factors are shown. *GT-1* shows similar expression to *DF1* in the tapetum while *GT-3a* is less highly expressed than DF1 but was also found to be tapetum-enriched relative to whole anthers.

LBD2 is part of the *ASYMMETRIC LEAVES2/LATERAL ORGAN BOUNDARIES* (*AS2/LOB*) family that contains 42 members [214]. The only other LBD-family transcription factor highly expressed in the tapetum is *LBD41.* While *LBD41* is not tapetum-specific like *LBD2* it is tapetum-expressed, and highly enriched (fold changes > 36) in the tapetum relative to meiocytes and leaves (Fig. 2-9). *LBD41* has been implicated in response to hypoxia [217], which has been shown to function in

maize anthers to control the specification of meiocytes [20]. *LBD41* may function in the tapetum at a similar developmental stage, while hypoxia is naturally generated in growing anthers [20].

2.6 DAP-seq datasets suggest novel transcriptional regulation in the tapetum

Transcription factors were investigated in the tapetum due to their ability to affect large numbers of downstream genes and regulate key processes. Regulatory dynamics between transcription factors also play important roles in controlling gene expression, cell state, and developmental progression. There are several transcription factors known to function in the tapetum, most well studied of these are the factors involved in the transcriptional cascade from DYT1 to MS1 (Fig. 2-10) [80].

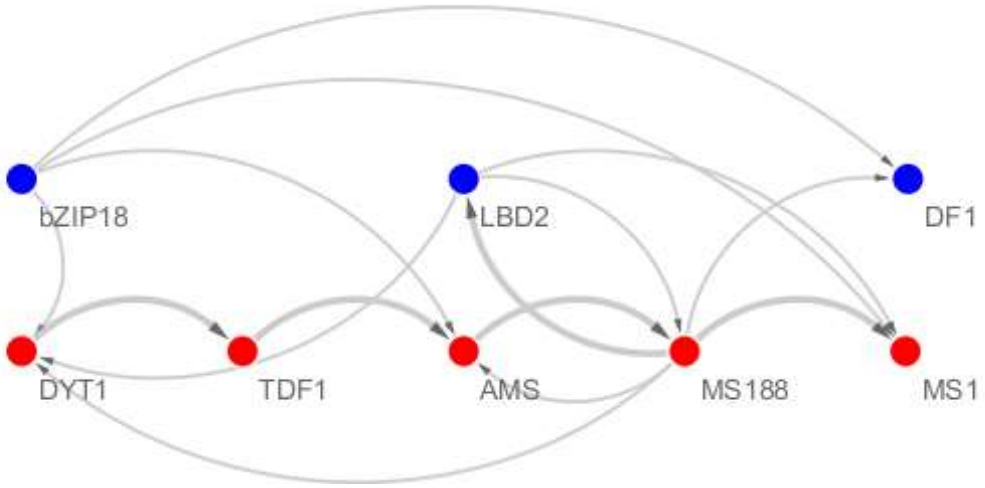


Figure 2-10: A transcriptional network of known tapetal transcription factors and the inferred interactions with selected tapetum-enriched transcription factors. Transcriptional interactions from literature are represented by bold arrows, whereas DAP-seq predicted interactions are shown by thinner arrows. The DYT1-TDF1-AMS-MS188-MS1 transcriptional cascade is well studied [80]. *LBD2* has been found to be downstream of MS188 [204] but this interaction was not predicted from DAP-seq data [218]. All downstream targets of bZIP18, LBD2 and DF1 are predicted and have not been experimentally verified. Novel regulatory interactions of MS188 are predicted for *DYT1*, *AMS*, and *DF1* from DAP-seq data. These predicted interactions between transcription factors could explain mutant phenotypes seen for selected transcription factor genes.

DNA affinity purification sequencing (DAP-seq) has allowed the identification of thousands of putative binding sites for 529 transcription factors, roughly one third of those found in *Arabidopsis thaliana* [218]. This data has facilitated the expansion of the

tapetal transcriptional network, with putative interactions between known transcription factors and those identified from the tapetum transcriptome.

The core pathway is predicted to progress from DYSFUNCTIONAL TAPETUM 1 (DYT1) to DEFECTIVE in TAPETAL DEVELOPMENT and FUNCTION 1 (TDF1), ABORTED MICROSPORES (AMS), MS188, and finally MS1. Predicted interactions of candidate transcription factors with core factors may hint at stage-specific expression and the tapetal processes regulated. From DAP-seq data, MS188 is predicted to be upstream of *DYT1* as well as *AMS* (Fig. 2-10). This interaction with *AMS* may represent a positive feedback loop as AMS and MS188 act together to regulate gene expression [91]. MS188 is predicted to regulate *DF1* (Fig. 2-10). Given the role of DF1 in regulating seed mucilage [203] and the tapetal (from stage 8), and pollen phenotypes of *df1*, suggests a late tapetal function in regulating pollen wall biosynthesis (Fig. 2-10) [42]. *LBD2* has previously been proposed as a downstream target of MS188 [204]. From DAP-seq data LBD2 is itself predicted to regulate *MS188*, possibly suggesting another feedback loop (Fig. 2-10). LBD2 and bZIP18 are both predicted to regulate DYT1 and MS1 (Fig. 2-10).

DAP-seq data provides candidate genes, that may be mis-regulated in mutants and so contribute to the tapetal phenotypes observed (Fig. 2-2). Predicting which genes may be responsible for the phenotypes is difficult, not least because the number of downstream target genes predicted from DAP-seq varies greatly. For the three selected transcription factors there is a large variation in the number of downstream target genes, from 472 for DF1, to 3382 for bZIP18, and 4465 for LBD2. GO term enrichment analysis was performed to find biological processes which may be contributing to tapetal phenotype [189-191].

As a proof of principle, the 3614 DAP-seq targets of MS188 were investigated [218]. There is a significant enrichment for GO terms relating to “phenylpropanoid metabolic process” (GO:0009698, $p = 5.08\text{x}10^{-5}$), “fatty acid metabolic process” (GO:0006631, $p = 3.57\text{x}10^{-4}$), and “external encapsulating structure organization”

(GO:0045229, $p = 3.25x10^{-4}$). This matches the known function of MS188 in regulating sporopollenin biosynthesis metabolic pathways [89, 91].

Targets of bZIP18 are significantly enriched for genes involved in "reproductive system development" (GO:0061458, $p = 4.80x10^{-4}$), including tapetum transcription factor genes *AMS*, *DYT1*, *MS1*, *bHLH10*, and *bHLH89*. bZIP18-targets are also enriched for genes functioning in "acyl-CoA metabolic process" (GO:0006637, $p = 1.42x10^{-4}$). This suggests that bZIP18 functions in regulating tapetal development, fatty acid metabolism, and sporopollenin biosynthesis, processes which span many stages of tapetal development.

For LBD2-targets, the most significantly enriched GO term is "transmembrane receptor protein tyrosine kinase signalling pathway" (GO:0007169, $p = 3.45x10^{-4}$), suggesting a role in regulating extracellular signalling between the tapetum and other cell types. Within this group of genes is *SCHENGEN 3* (*SGN3*), which functions in regulating casparian strip formation [219]. Similarities between casparian strip and pollen wall biosynthesis, as well as microspore expression of *SGN3*, prompted us to investigate if the SGN3-pathway could also function in pollen development. Pollen phenotype of a T-DNA insertion mutant of *SGN3* was investigated by SEM but did not differ significantly from WT ($p < 0.05$, T-test).

LBD2-targets are also enriched for genes associated with "organic hydroxy compound metabolic process" (GO:1901615, $p = 1.68x10^{-4}$), including genes functioning in the phenylpropanoid pathway, such as *PHENYLALANINE AMMONIA LYASE 1* (*PAL1)* (phenylpropanoid pathway [57, 59]) as well as a host of Cytochrome P450 enzymes. The mis-regulation of such genes, in the *lbd2* mutant, may underlie the pollen phenotypes observed (Figs. 2-7 & 2-8).

DF1 DAP-seq target genes were not significantly enriched for any associated GO terms. Microarray data from GFP-tagged DF1 expressing root hairs have been used previously to find DF1-activated and DF1-suppressed genes (638 activated and 819 suppressed) [202]. Both activated and suppressed genes were compared to targets

predicted by DAP-seq and no overlap was found with either list. Genes upregulated by DF1 [202] are significantly associated with response to environmental stimuli. Downregulated genes are associated with GO terms for "cell wall pectin metabolic process" (GO:0052546, p = 2.92x10^{-11}), suggesting that DF1 may act as a negative regulator of pectin and cell wall biosynthesis.

2.7 qRT-PCR does not reliably confirm target genes of tapetum-expressed transcription factors

While target genes have been predicted for candidate transcription factors, direct evidence of transcriptional regulation is still needed. To test these relationships, qRT-PCR was performed on cDNA from whole inflorescences of WT and mutant plants. Target genes that are greater than 2-fold enriched in the tapetum relative to whole anthers, and highly expressed in the tapetum were selected (Fig. 2-11) (section 2.3). Where possible, target genes that are regulated by multiple transcription factors were selected for testing.

The genes selected have been shown directly to function in the tapetum or pollen development (*ABCG1* [220, 221], *ACETOACETYL-COA THIOLASE 2* (*ACAT2*) [222], *FLAKY POLLEN* 1 (*FKP1*) [223], *LESS ADHERENT POLLEN 3* (*LAP3*) [60], and *DETOXIFYING EFFLUX CARRIER 35* (*DTX35*) [224]) or function in processes that occur in the tapetum (*AT2G21430* [225], *ARABIDOPSIS THALIANA METACASPASE 1* (*ATMC1*) [226], *MEVALONATE 5-DIPHOSPHATE DECARBOXYLASE 1* (*MVD1*) [227], and *NRPB6B/NRPE6B* [228]). As such, the mis-regulation of these genes may underlie the tapetum and pollen phenotypes observed (Figs. 2-2, 2-7 & 2-8).

Expression was compared to the *ACTIN8* (*ACT8*) housekeeping gene for three biological replicates of each genotype, repeated six times. None of the transcription factors are predicted regulators of *ACT8*, so it is not expected to be variably expressed between genotypes [218]. WT and mutant expression was compared for individual genes by ANOVA at the ΔCt level. Most genes (7 out of 9) were not significantly differentially expressed between WT and any mutant (Fig. 2-12). *ABCG1* and *DTX35*

were significantly more highly expressed in *df1* than WT ($p < 0.001$ and $p < 0.01$ respectively, ANOVA) (Fig. 2-12). In *lbd2*, *ABCG1* was significantly more highly expressed than in WT ($p < 0.05$, ANOVA) (Fig. 2-12). No genes were significantly differentially expressed between WT and *bzip18*.

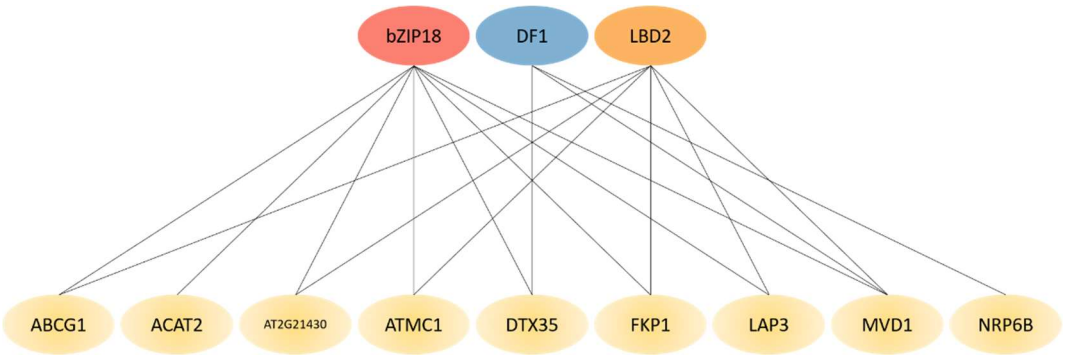


Figure 2-11: Downstream targets, predicted from DAP-seq [218], of bZIP18, DF1, and LBD2 tested by qRT-PCR. All targets tested are enriched in the tapetum relative to whole anthers by greater than 4-fold.

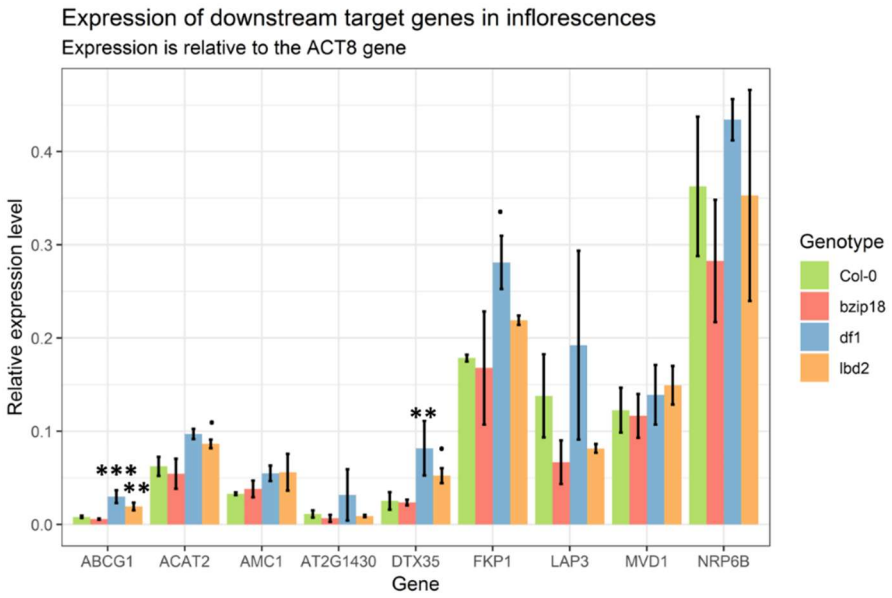


Figure 2-12: Relative gene expression of tested target genes in WT, *bzip18*, *df1*, and *lbd2* inflorescences. Expression is normalised to the ACTIN8 housekeeping gene. Error bars represent 95% confidence intervals from three biological replicates. The expression of individual genes was compared between mutant and WT plants by ANOVA. Significance values represent comparisons between mutant and WT plants; "." = 0.1, "*"=0.05, "**"=0.01, "***"=0.001. Error bars represent 95% confidence intervals.

This confirms the proposed links between DF1 and *DTX35*, as well as LBD2 and *ABCG1*, but also, suggests that DF1 regulates *ABCG1*, which was not predicted by

DAP-seq (Fig. 2-12). All significantly different genes are more highly expressed in mutants, suggesting the transcription factors are negative regulators of these targets in WT. It is hard to reconcile negative regulation of factors required for pollen development with the pollen phenotypes of *lbd2* and *df1*, though it is interesting to note that both *ABCG1* and *DTX35* (also known as *FLORAL FLAVONOID TRANSPORTER* (*FFT*)) are both transporters. As LBD2 regulates *PAL1*, which catalyses the deamination of phenylalanine for flavonoid and lignin biosynthesis, negative regulation of export could act to control pathway flux and the build-up of intermediates for further chemical modification.

As well as genes, the expression of transposons varies between cell types and has particularly important consequences in the sexual lineage. Transposon expression can be indicative of a particular chromatin environment or function performed by a cell. To discover similarities to other cell types, particularly those of the sexual lineage, transposon expression was analysed in the tapetum.

2.8 The tapetum does not show higher levels of transposon expression than other somatic tissues

Developmental relaxation of transposon silencing occurs at specific stages through the *Arabidopsis* life cycle, including in meiocytes and pollen. It is known that transposon expression in the vegetative cell gives rise to sRNAs to reinforce silencing at transposons in the sperm [105]. To investigate if the tapetum also participates in developmental relaxation of transposon silencing, I produced total-RNA sequencing libraries of FACS-sorted tapetal cells. Libraries from three biological replicates were sequenced with a Nextseq 500 to a mean depth of 27.5M reads.

Transposon expression in the tapetum was compared to rosette leaf, meiocyte, and pollen, using Kallisto and Sleuth [229, 230]. Of the tissues tested, the tapetum expresses the lowest number of transposons, 1380. This compares to 1390 detected in leaf, 2603 in pollen and 5121 detected in meiocytes (transcripts per million (TPM) > 0) (Fig. 2-13). Most transposons are very lowly expressed, so a cut-off of four TPM was

applied to investigate highly expressed transposons. 180 transposons were expressed above this level in the tapetum, 156 in the pollen, 261 in meiocytes, and 220 in leaves (Fig. 2-13).

To investigate similarities between the tapetum and other tissues in terms of transposon expression, the expression of all detected transposons were compared. Transposon expression in the tapetum correlates best with leaf and meiocyte (R^2 = 0.672 and 0.627 respectively, Pearson's product-moment correlation) but shows much lower correlation with pollen (R^2 = 0.395) (Fig. 2-14). Size distributions of well expressed transposons were compared between tissues but were found to not be signifcantly different ($p > 0.05$, ANOVA). The tapetum, therefore, appears to not differ significantly from leaf in terms of the number and size of transposons expressed.

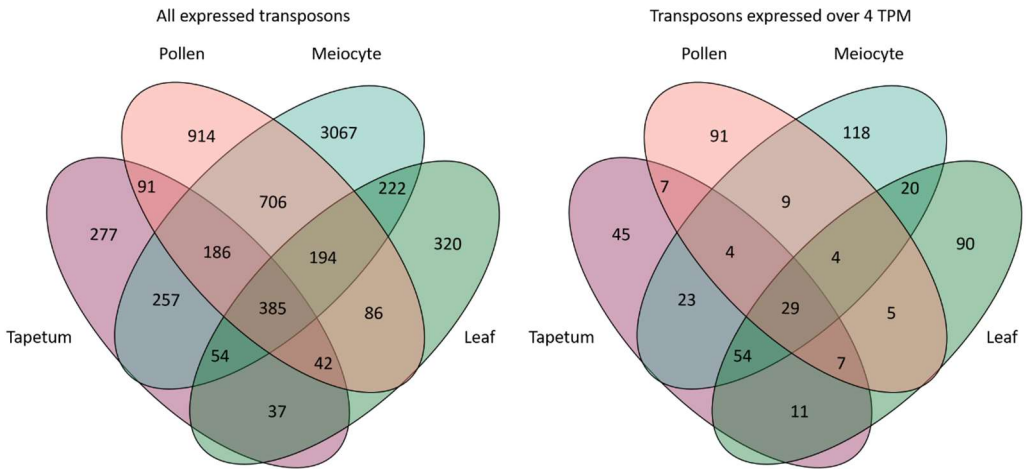


Figure 2-13: Venn diagrams representing common and uniquely expressed transposons between tapetum, pollen, meiocytes, and rosette leaf. Transposons that are detected at any level, and those expressed at greater than four transcripts per million (TPM) are represented.

206 transposons were significantly differentially expressed between the four tissues ($p < 0.05$, Sleuth), and of these 32 had their highest level of expression in the tapetum. However, none of these transposons were expressed uniquely in the tapetum. All 32 significantly tapetum-enriched transposons were expressed at greater than 4 TPM in at least one other tissue.

While the tapetum does not appear to uniquely express transposons, I wanted to investigate if highly expressed transposons were derived from specific transposon

superfamilies. The 180 highly tapetum expressed transposons were classified into superfamilies and compared to each family's genomic abundance (Fig. 2-15). The proportions of each superfamily expressed in the tapetum matches closely that seen in the genome. Only RC/*Helitron* family transposons are significantly under-represented ($p < 0.01$, χ^2 test), and unassigned transposons are significantly enriched ($p < 0.001$, χ^2 test) (Fig. 2-15).

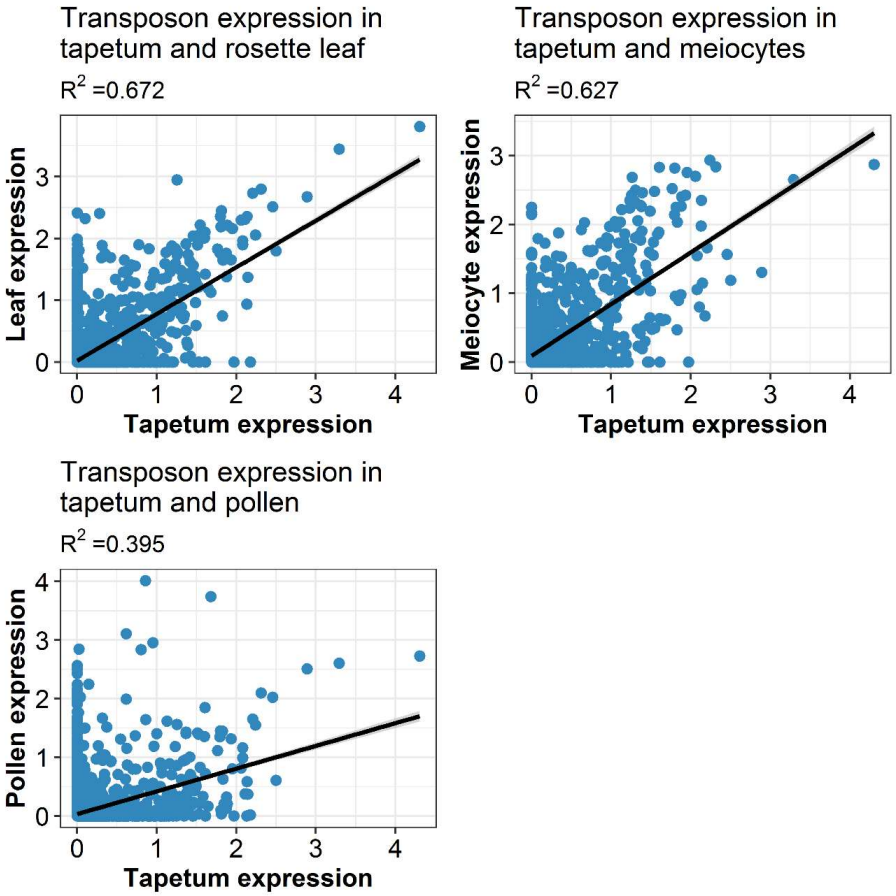


Figure 2-14: Transposon expression in the tapetum compared to leaf, meiocyte, or pollen. When the expression of all detected transposons is compared in different tissues good correlations are found between the tapetum and leaf, and tapetum and meiocytes. Expression in pollen correlates less well, suggesting that pollen expresses a more unique suite of transposons. Tapetum transposon expression does not appear to differ significantly, in terms of number or composition, from leaf.

In total there are 22 unassigned transposons expressed in the tapetum, which includes the *Sadhu* family [231]. I investigated if the enrichment for unassigned transposons was tapetum specific and found that 21 were also expressed in at least one other tissue at greater than 4 TPM, and one was expressed in other tissues below

this level. Unassigned/*Sadhu* transposons are therefore generally well expressed across a variety of tissues, despite their small number in the genome.

Together, these data suggest there is little tapetum-specific transposon expression, and the transposon families enriched in the tapetum are generally expressed in other tissues. In terms of number, size, and level of expression, the tapetum does not differ markedly from other somatic tissues such as rosette leaf. It appears, therefore, that the tapetum is not a site of transposon activity, differing from the meiocytes and pollen vegetative cell.

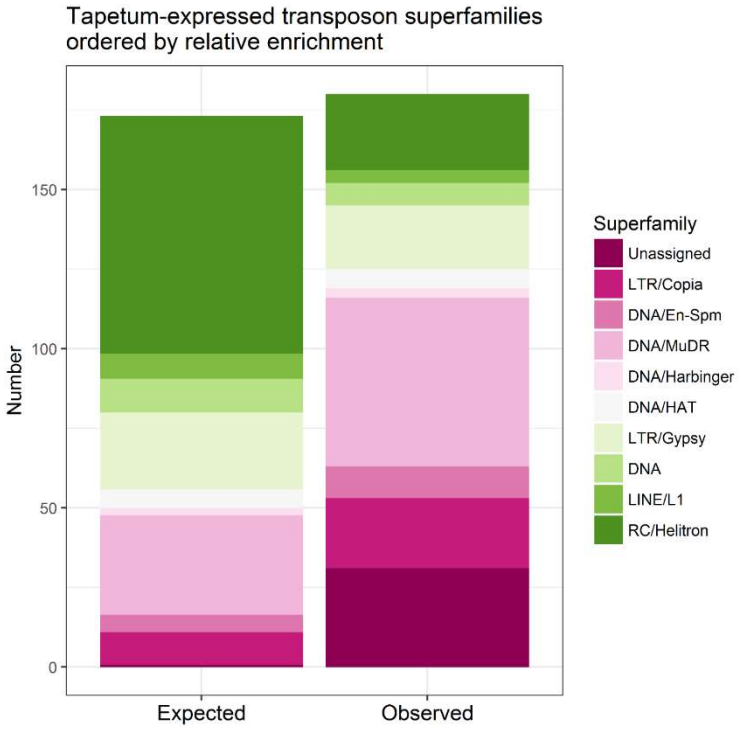


Figure 2-15: Superfamily composition of tapetum transposons expressed at greater than four transcripts per million. Expected represents the number of transposons of each superfamily expressed based purely on their abundance in the genome. RC/Helitron elements are significantly under represented compared to their abundance in the genome, and Unassigned elements are significantly over represented. Unassigned elements, including the *Sadhu* family, appear expressed at high levels in all tissues tested. Superfamilies not present in the observed data are not significantly different from their expected frequenices.

2.9 Discussion

While the tapetum has been an important target in the study of plant fertility, until recently research has been hindered by the lack of available transcriptomic data. RNA-seq data has provided confirmation of the expression of genes known to function in the tapetum and the identification of novel tapetum-expressed genes. FACS-sorted cells were found to be of high purity and have yielded high quality RNA-seq libraries, that compares well to published data from *Arabidopsis* [85].

2.9.1 Novel genes regulate both tapetal and germline development

Analysis comparing the tapetum transcriptome to other plant tissues was undertaken. This enabled the identification of novel tapetum-enriched and tapetum-specific genes. Knowledge of genes preferentially and specifically expressed in the tapetum is an important contribution to tapetal research. Tapetum-specific genes may also be a useful tool in the creation of transgenic model and crop plants, with tapetum-expression being tuned to different levels or to specific stages. Understanding the full suite of genes expressed in the tapetum will aid the elucidation of tapetum transcriptional networks, signalling, and metabolic pathways.

It was shown that the tapetum-enriched gene *DF1*, and the tapetum-specific gene *LBD2* regulate tapetal and pollen development. Both the *df1* and *lbd2* mutants show tapetal defects from stage 8 and an increased rate of multi-layered tapetum. Defects were particularly pronounced in *lbd2* anthers at early stage 9 which showed a significant rate (6.5%) of highly vacuolated tapetum (Fig. 2-2). At late stage 9 a significant proportion of locules showed expanded and early degenerating tapetum (11.2%). An increased rate of misshapen pollen is also seen in both lines, though this is not associated with an increased rate of aborted pollen from Alexander staining. In *lbd2*, this increased rate of misshapen pollen was associated with rare anthers with very high rates of defective pollen, which also showed defects in pollen exine patterning. This was also apparent to a lesser extent in *df1*.

More work is required to fully characterise the roles of these transcription factors in the tapetum. While I have shown that both *df1* and *lbd2* show higher rates of misshapen pollen than WT, whether this causes the reduced fertility (Fig. 2-4) has not been shown directly. Reciprocal crosses between WT and mutant plants would confirm the male fertility defects of these mutants. Pollen germination assays could also be performed to test if misshapen pollen are still able to grow and fertilise egg cells. Mutant phenotypes must be confirmed with other alleles (Fig. 2-3), particularly full knock-outs which could be generated through CRISPR editing. It must be tested whether the phenotype can be complemented with the reintroduction of the native gene. Expression through developmental time could be followed through *in situ* hybridisation, or the creation of GUS or fluorescent reporter constructs. I have identified other transcription factor family members expressed in the tapetum which could be acting redundantly with the selected transcription factors. Double mutant plants would confirm if there is genetic redundancy between family members.

From expression [204] and DAP-seq [218] data, *LBD2* and *DF1* are predicted to be downstream targets of MS188, which suggests expression after anther stage 6 [80]. Tapetal and pollen phenotypes are compatible with this expression pattern, as defects are not visible before stage 8. Analysis of transcription factor binding sites by DAP-seq has provided candidate target genes for 529 *Arabidopsis* transcription factors [218]. This data has provided target genes and biological processes which may explain the mutant phenotypes observed. DAP-seq predicted targets support a role of DF1 and LBD2 in regulating sporopollenin biosynthesis, which were also inferred from pollen phenotypes (Fig. 2-8). Testing downstream targets of *bZIP18, DF1,* and *LBD2* that function in the tapetum showed very few target genes were differentially expressed between mutant and WT inflorescences (Fig. 2-12). The two genes significantly more highly expressed in *lbd2* than WT, *ABCG1* and *DTX35*/*FFT*, are exporters of sporopollenin precursors and flavonoids respectively [221, 224]. LBD2 also regulates an entry point to these pathways, *PAL1*. It is possible that LBD2 acts as a negative regulator of export to control flux through these pathways and increase the concentration of intermediates, which are extensively modified in the tapetum [42].

Differential expression of non-predicted target genes was also observed, suggesting putatively indirect regulation.

The lack of mis-expression of many target genes could be explained by several reasons. Firstly, genes that are targets of multiple transcription factors were chosen to test more interactions in one experiment. This may lead to target genes not being differentially expressed due to redundancy. Secondly, as the transcription factors and target genes are all enriched in the tapetum, the use of whole inflorescence cDNA may have obscured real changes in gene expression. This could be due to changes in the tapetum being too subtle to detect, or the regulation of target genes in other cell types obscuring tapetal gene expression. Finally, as the *lbd2,* and *df1* plants show tapetal defects, such as multi-layered tapetum, the proportion of tapetal cells within whole anthers may not be constant between genotypes. qRT-PCR will be repeated with new biological replicates and cDNA. In future, gene expression of mutants can be analysed through RNA-seq of FACS-isolated tapetal cells to discover the processes disturbed in the *lbd2* and *df1* mutants. Direct targets could be confirmed by ChIP-seq to find promoter sequences bound *in vivo*.

2.9.2 The tapetum is not a hotspot of transposon expression

Meiocytes are a site of enhanced transposon expression [105], but RNA-seq data suggests that this phenomenon does not extend to the tapetal nurse cells. While the meiocytes and pollen express many transposons, the tapetum expresses a similar number to rosette leaf tissue, and expression of individual transposons also correlates best with the expression level in leaf. Few transposons are expressed uniquely in the tapetum and none are significantly so. Altogether, this would suggest that the tapetum is not a site of transposon activation, like meiocytes and the pollen vegetative cell, and that transposon expression is not employed to reinforce silencing in the meiocytes.

2.10 Summary

Access to high purity transcriptomes of the tapetum is a great advance in the study of tapetum function and its role in regulating germline development. While I have focused on the discovery of new transcription factors functioning in the tapetum, this data could facilitate the discovery of novel genes involved in any aspect of tapetum biology. As such, the data in this chapter have underpinned much of the work performed in other chapters. While many genetic screens have been performed to discover genes controlling male fertility, the approach employed in this chapter shows there are many smaller effect genes yet to be discovered that affect tapetum function and fertility.

2.11 Methods

2.11.1 Plant materials

All *A. thaliana* plants used in this chapter are of Col-0 ecotype, and grown under 16-h light/8-h dark in a growth chamber (21 °C, 70% humidity). T-DNA insertion mutants [232] were obtained from the Nottingham Arabidopsis stock centre (NASC) (http://arabidopsis.info/) and confirmed homozygous by genotyping with three primer PCR, with primers for the left T-DNA border (primers section 6). The *df1-1* (SALK_106258) line has been described previously [202]. Confirmed homozygous mutants and transgenics were sown directly onto soil without stratification.

2.11.2 Construction of *pA9::NTF*

The 979 bp region upstream of the *A9* TSS was amplified from genomic DNA using primers pHG001, and pHG002 (primers section 6). The *pA9* fragment was cloned into the pDONR207 Gateway donor vector. This was then combined with the nuclear targeting fusion (NTF) gene in a pMDC107 destination vector in an LR reaction [187]. *pA9::NTF* constructs were transferred to *Agrobacterium tumefaciens GV3101* for transformation into WT plants by floral dip [233]. Transformed seed were surface

sterilised and sown onto full strength Murashige & Skoog media (no glucose) containing hygromycin (25 ng/mL). Seeds were stratified at 4 °C for two days and then grown for one week with 16-h light 8-h dark at 20 ^{0}C. Positive transformants were transferred to soil. Plants with strong and specific GFP expression were selected using a Zeiss 780 confocal microscope (Zeiss, Cambridge UK).

2.11.3 Protoplasting and cell sorting

Enzyme solution was created containing 0.4 M mannitol, 20 mM KCl, 10 mM $CaCl_2$, 20 mM MES (pH 5.7), 0.1% BSA (w/v), 1.25% Cellulase (w/v) (Yakult, R10), 0.1% Pectolyase (w/v) (Yakult, Y23), 1% Hemicellulase (w/v) (Sigma Aldrich, H2125), 2% Pectinase (w/v) (Sigma Aldrich, P2401), and 2% Beta glucanase (w/v) (Sigma Aldrich, G4423). Sterile distilled water, Mannitol, MES, and KCl were first combined and heated to 70 ^{0}C for five minutes. Enzymes were then added and dissolved by vortexing. The solution was then stored at 55 ^{0}C for 10 min, then cooled on ice for 10-15 min. $CaCl_2$ and 30% BSA (w/v) were then added, the solution was mixed thoroughly and stored at room temperature for digestion.

W5 solution was made with 154 mM NaCl, 125 mM $CaCl_2$, 5 mM KCl, and 2 mM MES (pH 5.7). W5 was filter sterilised through a 0.22 μm membrane (Sartorius, 16532). EW solution was made with 0.2 M mannitol, 67.5 mM $CaCl_2$, 50 mM NaCl, 12.5 mM KCl, 11 mM MES (pH5.7), and 0.05% BSA (w/v). EW solution was passed through a 0.45 μm filter (Sartorius, 16555).

Healthy inflorescences were picked, and open flowers and the largest 4 unopened buds removed to reduce pollen content. Inflorescences were placed on a microscope slide covered with double-sided tape and cut well (roughly eight times) with a sharp razor blade. Forceps and a scalpel were used to place the chopped material into a 70 μm filter (Falcon, 352350) in a small petri dish, filled with 6 mL of enzyme solution. Razor blades were changed regularly. Petri dishes were covered and incubated on a thermal block set at 23 ^{0}C. Solutions were incubated for five hours from the mid-point of inflorescence chopping. After incubation, the enzyme solution was gently collected

into 15 mL tubes with Pasteur pipettes. The same volume of W5 solution was added to each tube and mixed by gentle pipetting. A small amount of un-digested material was added to the tubes to aid sedimentation of protoplasts. Solutions were centrifuged at 190 g for 5 min at 21^{0}C, with the acceleration and brake set low (setting 2 of 9). The supernatant was removed, and the pellet gently resuspended in 1.5 mL of EW solution. The volume can be varied but it is important that the protoplast solution is not too dense for sorting. A small volume of protoplast solution was checked under a light microscope for protoplast health before passing through a 30 µm filter (Sysmex, 04-0042-2316).

Protoplasts were sorted on a Sony SH-800 using a 100 µm chip (Sony Biotechnology, Pencoed, UK). After doublet discrimination, GFP-positive protoplasts were identified by their separation from non-fluorescent cells along a GFP axis when fluorescence was plotted against forward scatter. Tapetal cells showed roughly 1000-fold higher fluorescence than non-tapetal cells. GFP fluorescence in anthers and protoplasts from *pA9::NTF* were imaged on a Leica DM6000 (Leica, Milton Keynes UK).

2.11.4 RNA-sequencing library production

For mRNA libraries, batches of 400 tapetal protoplasts were sorted into 1.5 mL eppendorf tubes containing 2x RNase-free sodium lysis buffer; 200 mM Tris-HCl (pH7.5), 1 M NaCl, 20 mM EDTA (pH 8), 10 mM DTT, and 2% (v/v) SDS. Tubes were centrifuged at >10,000 g for 5 minutes and stored at -70 ^{0}C. RNA-seq libraries were created according to [188].

For total RNA, libraries between 6500 and 8500 cells were collected into *mir*Vana lysis buffer (ThermoFisher, AM1560) and total RNA was extracted according to the manufacturer's instructions. RNA-sequencing libraries were created as in [9].

2.11.5 RNA-seq mapping and gene expression analysis

RNA-seq data from mRNA libraries were mapped to the TAIR10 genome with Tophat and differentially expressed genes found with Cuffdiff [234]. Tapetum-

specific genes were found using the R package CummeRbund [211]. For differential gene expression analysis, the tapetum was compared to whole anther RNA-seq data from stage 4-7 WT anthers [82]. For the specificity analysis the tapetum was compared to data from; stem [235], seedling [236], embryo [237], rosette leaf [126], and root epidermis, cortex, endodermis, stele, and columella [151], as well as meiocyte and pollen (unpublished data).

2.11.6 Inflorescence sectioning

Sectioning of *Arabidopsis* inflorescences was performed as in [238] with 4% paraformaldehyde in PBS as fixative. Resin blocks were cut into 5 μm thick sections.

2.11.7 Alexander staining

Alexander staining was performed as in [213].

2.11.8 Silique length measurements

The bottom ten fully expanded siliques from the main inflorescence of WT and mutant plants were collected from five individuals. Siliques were photographed and measured electronically using Image J [239, 240].

2.11.9 Scanning electron microscopy and pollen counting

Scanning electron microscopy was performed on a Zeiss Supra 55 VP FEG (Zeiss, Cambridge UK). Two individual anthers each from nine plants were placed onto imaging platforms before being gold coated. All samples were imaged at a power of 3 kV and a magnification of 1000, or 12000 times. SEM images were anonymised, and the healthy and misshapen pollen counted blind using the Cell Counter plugin in Image J (https://imagej.nih.gov/ij/plugins/cell-counter.html) [239, 240]. Anthers with the highest and lowest rates of misshapen pollen were removed for statistical analysis, giving n = 16.

2.11.10 qRTPCR

Whole inflorescences were collected in the same manner as for protoplasting and cell sorting (2.11.3). Four inflorescences each from three individual plants were collected for each biological replicate. Three replicates from each genotype were collected and RNA was extracted using TRIzol (Sigma Aldrich, T9424) according to manufacturer's protocol. RNA was DNase-treated using a Turbo DNA-free™ kit (Thermo Fisher scientific, AM1907) according to manufacturer's instructions. 917 ng total RNA from each sample was reverse transcribed with a RevertAid First Strand cDNA Synthesis Kit (Thermo Fisher scientific, K1621). For mutant expression analysis, cDNA was created using only Oligo dT primer, and for downstream target expression both Oligo dT and random primers were used in a 1:1 mix. qRT–PCR was performed with SYBR Green (Roche, 4707516001) on a LightCycler 480 Real-Time PCR System (Roche). Mutant expression analysis was performed with three technical replicates on one plate and downstream target analysis with six technical replicates over two plates. Ct values were averaged within plates for biological replicates and ΔCt values (The number of cycles between the gene of interest and housekeeping crossing the threshold level) were averaged across plates after comparison to the reference gene *ACTIN 8*. Primers are listed in section 6. The expression of genes was compared between genotypes by ANOVA at the ΔCt level. Data are presented as relative expression to ACTIN8 ($2^{-\Delta Ct}$) (target genes), or as expression relative to WT (mutant expression).

2.11.11 Transcriptional network

Transcriptional interactions were collected from literature or from DAP-seq datasets [218]. If multiple lists of targets were available for transcription factors then genes common to both lists were taken as target genes. Transcriptional networks were created using Cytoscape [241].

2.11.12 Transposon expression analysis

Total RNA-seq libraries were aligned to *Arabidopsis thaliana* TAIR10 transposon and gene annotations using the fast Kallisto pseudo-aligner [230]. Tests for significant differential expression were performed within Sleuth [229]. Three tapetal total RNA libraries were compared to three WT meiocyte replicates [9], three replicates of leaf, and four replicates of pollen (unpublished data).

2.11.13 Plotting

All violin, bar, scatter, and tile (expression matrices) plots were created using ggplot2 [242]. Boxplots represent the data within the first and third quartile, with the darker line representing the median. Whiskers extend to 1.5x the inter-quartile range, or to the furthest data point. For mutant expression qRT-PCR error bars on bar charts represent standard error, and for downstream target qRT-PCR represent 95% confidence intervals.

2.11.14 GO enrichment and statistics

GO term enrichment was calculated with Panther [191] *p*-values reported are from Fisher's exact test with false discovery rate multiple test correction at $p < 0.05$.

For comparisons of non-normally distributed data a Kruskal-Wallis test was performed. To find pairwise differences between samples *post-hoc* analysis was performed using the Conover-Iman test and implementing the Holm correction to control false discovery rate ($p < 0.05$).

Chapter 3 Single-cell transcriptomes of the tapetum reveal stage-specific gene expression

3.1 Introduction

Cells are a fundamental unit of biological systems, from bacterial communities to complex multicellular organs. Understanding the differences between individual cells that constitute these systems is essential to gaining a deeper understanding of biological function. The ability to assay the genome, epigenome, transcriptome, proteome, and metabolome of single cells is a revolutionary advance in biology and has provided new insights in both applied and fundamental research [243].

But why should one sequence single cells when bulk sequencing already provides so many answers to biological questions? Firstly, we would expect some level of variation within any biological system, meaning that a bulk transcriptome cannot be representative of every cell within the system. Secondly, there is growing evidence that the behaviour of individual cells in a system do not necessarily reflect quantitative changes observed for the population as a whole. For example, polycomb silencing and gene expression at the FLC locus in *Arabidopsis* occurs in a digital manner in individual cells but appears as a gradual change at the organ level [244, 245].

Investigating transcription at a bulk level may not just be a less accurate representation of the population, but may be actively misleading [246]. As a hypothetical example, if we were to follow a population of cells through a time course experiment and found that two genes, *A* and *B*, were induced at the same timepoint we would naturally assume they are regulated together and may function in the same pathway (Fig. 3-1). If we performed the same experiments but sequenced single cells we may find that *A* and *B* are induced at the same time but are expressed in different

sub-populations of cells, and never co-expressed (Fig. 3-1). From these data we would arrive at very different conclusions about the functions of *A* and *B*. Indeed, considering groupings of cells may alter the relationships between genes which have been inferred from population-level analyses [246].

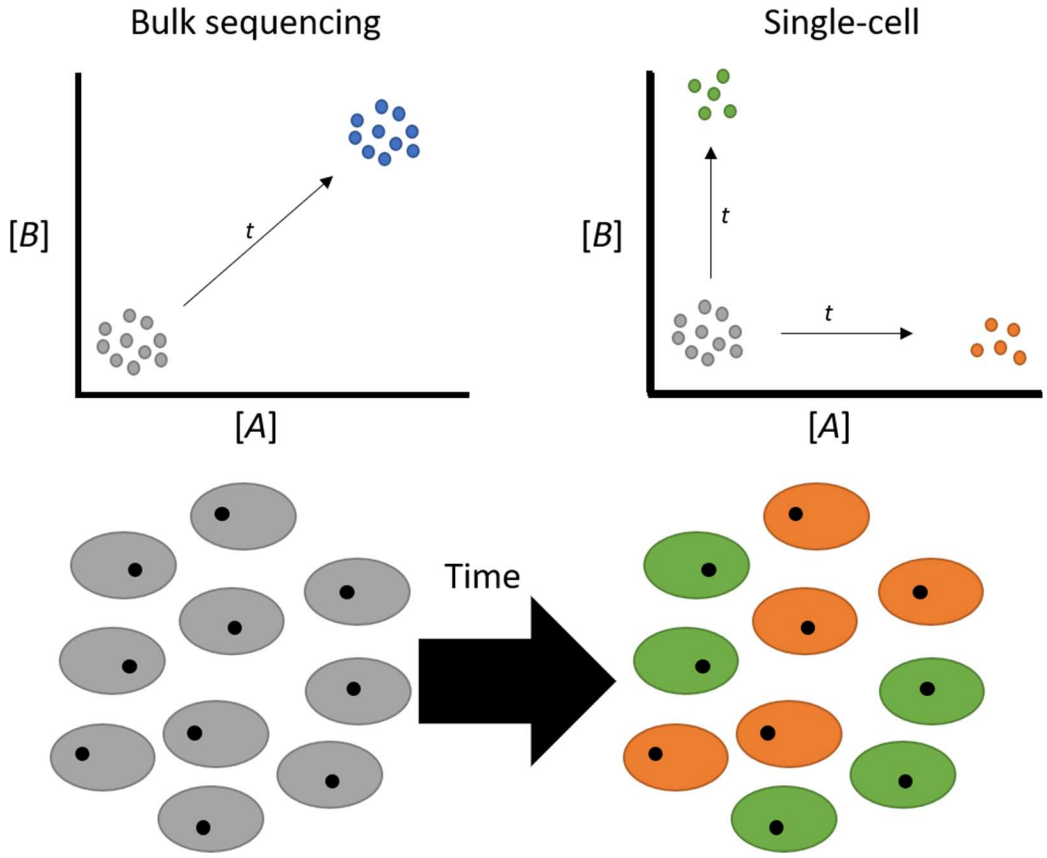


Figure 3-1: Analysis of gene expression in a population of cells over time can yield different results if bulk sequencing or single-cell sequencing is employed. Genes *A* (orange) and *B* (green) in a population of cells after a time, t, in sub-populations of cells. When cells are analysed in bulk, A and B are found to be induced together and are wrongly assumed to be co-expressed in all cells (blue cells). Only by sequencing single cells can it be seen that expression of *A* and *B* is mutually exclusive and occurs in distinct sub-populations of cells.

Single-cell sequencing requires the precise extraction of individual cells of interest, with a minimal chance of collecting multiple cells. Manual extraction of cells is too low throughput to study many aspects of single-cell biology but is well suited to studies that require tracking of cell divisions, such as in embryogenesis [247]. Fluorescence-activated cell sorting (FACS)-based approaches are broadly applicable to many single-cell experiments and can easily provide hundreds of single cells for experiments. Demand for experiments that use thousands, rather than hundreds of

cells, has pushed the use of microfluidics for isolation. In these approaches, cells are captured in individual droplets or nanowells, maximising throughput and minimising cost [248]. Much of single-cell research has focused on mammalian cell cultures, blood cells, and cancers due to their ease of isolation. While plant cells are more difficult to isolate from rigid cell walls, most of the techniques discussed in this chapter are also broadly applicable to plant research.

Since the first experiment describing RNA-sequencing of a single cell [249] the techniques available to assay single-cell 'omics data have increased in both their breadth and ease of use. Single-cell experiments are fast becoming a readily available tool for most researchers. Techniques such as Smart-seq2 [188] massively reduced the cost of mRNA library preparation, allowing tens to hundreds of cells to be assayed in the same experiment. The single-cell isolation and barcoding steps involved in plate-based assays precludes most experiments with greater than 1000 cells. With techniques such as Drop-seq [250, 251] and the 10x Genomics platform [252] thousands to hundreds of thousands of cells can be assayed in one experiment, though usually to a lower depth per cell. A recent paper has applied microfluidics to profile >4000 protoplasts from the *Arabidopsis* root, identifying multiple cell types and gene expression changes [251]. The number of aspects of cell phenotype that can be assayed in single cells is ever increasing, with DNA and RNA the easiest to analyse through the sequencing library methods mentioned previously.

The genome is commonly assumed to be stable throughout the life cycle, but mutations can occur in somatic and germline cells, causing pathologies such as cancer in animals and mutation which can be inherited through germlines. This genomic heterogeneity can be investigated through sequencing of single cells. Due to the need to amplify DNA before library production, and the random drop-out and errors caused by amplification bias, the analysis of single nucleotide variants is extremely challenging. Single-cell experiments have therefore focussed on larger differences, such as copy number variation and large chromosomal rearrangements [253, 254].

As well as the sequence of the genome, we can also investigate covalent modifications, protein association, and chromatin structure of the genome at the single-cell level. DNA methylation of cytosine is a common epigenetic modification in eukaryotes and can show cell-type-specific distributions across the genome [9, 151]. Single-cell bisulphite-sequencing (scBS-seq) facilitates the analysis of DNA methylation variation between single cells but, due to the destructive nature of bisulphite treatment, coverage of data per cell is low [255]. To overcome this problem, scBS-seq analysis requires the comparison of populations of single cells, or imputation of missing data [256]. Chromatin accessibility is often associated with promoters of actively expressed gene and enhancer elements. Changes in chromatin accessibility are associated with different cell types and responses to stimuli. At the single-cell level, chromatin accessibility can be analysed through digestion with DNase I, which digests open chromatin [257, 258]. An alternative is the ATAC-seq protocol, in which cells or nuclei are incubated with the T5 transposase which fragments DNA in open chromatin and ligates adapters in the same step [259, 260].

As well as an increasing number of single-cell 'omics techniques, there have been huge advances in our ability to assay multiple aspects of cell phenotype from single cells through single-cell multi-omics (reviewed in [261]). Genome and transcriptome sequencing from single cells (G&T-seq) [262] allows separation and sequencing of mRNA and DNA to investigate genomic and transcriptomic heterogeneity from the same population of cells. This has been expanded to also include analysis of single-cell DNA methylomes and transcriptomes (scM&T-seq) [263], linking dynamic changes in methylation state to transcription.

While the arsenal of tools to assay single-cell phenotypes continues to grow, the most common method employed is single-cell RNA-sequencing. The transcriptome is accessible and is highly dynamic, reflecting both large phenotypic differences in cell type and subtle/transient differences in cell state. It is this ability to identify even subtle differences between many single-cell replicates that has made single-cell transcriptome profiling such a powerful tool.

A common experimental approach with transcriptome data, with a large group of cells from a mixed population, is to classify them into cell types. This approach has been used to find sub-types of known cells, as well as new rare cell types such as hematopoietic stem cells [264, 265]. Searching for rare cell types is made easier by assaying a greater number of cells, making Drop-seq and 10X Genomics popular approaches for these experiments [250, 252]. Fortunately, cells can be clustered into groups from as few as 50,000-100,000 reads per cell. In all such experiments the transcriptome must be reduced in dimensionality to allow easier analysis and clustering. This is often performed with principal component analysis (PCA), or t-distributed stochastic neighbour embedding (t-SNE). From this dimensionally reduced data, cells can be clustered by a variety of methods (reviewed in [266]). Whichever method is employed, analysis of clustered data can be open to subjectivity and complicated by differences between cell type and cell state.

There is no comprehensive definition of cell type, and the criteria used to define types differs between biological systems. In the hematopoietic system, cell types are defined by a combination of renewal ability in *in vitro* and *in vivo* assays and expression of cell-surface markers. Within each type, cells can occupy one of several states such as quiescent, active cycling, and senescent [243]. In plant systems, it is common for cell type to be defined positionally within an organ, relative to other cell types or along gradients [1]. In the *Arabidopsis thaliana* anther, cell type is defined in a lineage and positional manner, and cells states are stages defined by physiological changes [10, 14].

Another approach to single-cell data is to represent cells as a continuum along a developmental process and to extract changes in gene expression along an inferred axis, known as pseudotime. While in many ways similar to time course experiments, pseudotime experiments have the added ability to deal with asynchronous development [267]. Lagging/quiescent cells can be separated from actively differentiating cells and so a more accurate picture of the development process in question can be built. A popular analogy, proposed by C. H. Waddington, is the

developmental landscape [268]. Like marbles rolling down a hill, cells move through the developmental topology and settle at their final cell fate. Single-cell sequencing allows cells to be visualised within a representation of this landscape [246]. The discovery of rare cell states and types by single-cell sequencing can therefore provide a deeper understanding of differentiation and developmental processes.

Pseudotime analysis builds on clustering of cells, as it attempts to find connections between cells in a dimensionally reduced representation of single-cell transcriptomes. Currently there is no best practice when it comes to analysing pseudotime experiments, and the approach used will differ based on the data collected and the hypothesis to be tested. As such, there is a wealth of pseudotime software packages available for the analysis of single-cell data.

In this chapter I have sorted single tapetal cells and sequenced their transcriptomes. I have taken three different approaches to analysing single-cell transcriptomes to recreate gene expression over tapetal development. From satisfactory pseudotime axes, I have been able to recreate expression of known tapetal genes, as well as discover novel tapetal gene expression profiles. I present evidence of a developmental and transcriptional switch occurring in the tapetum, as well as variable transcriptional dynamics. This shows both the feasibility of plant single-cell sequencing and the value of this work in discovering new genes and functions affecting tapetal and sexual lineage development.

3.2 FACS-sorted single tapetal cells yield good quality mRNA-seq libraries

Single tapetal cells were protoplasted and sorted from plants expressing a fluorescent nuclear-targeting fusion protein specifically in the tapetum, *pA9::NTF*, (2.11.3) [187]. Single cells were sorted into 96-well plates and mRNA libraries created using a modified Smart-seq2 protocol [188] from [262], with one well left blank and one well with 20 cells as controls (Fig. 3-2). All 20-cell wells, and most single-cell wells produced good quality cDNA for library preparation.

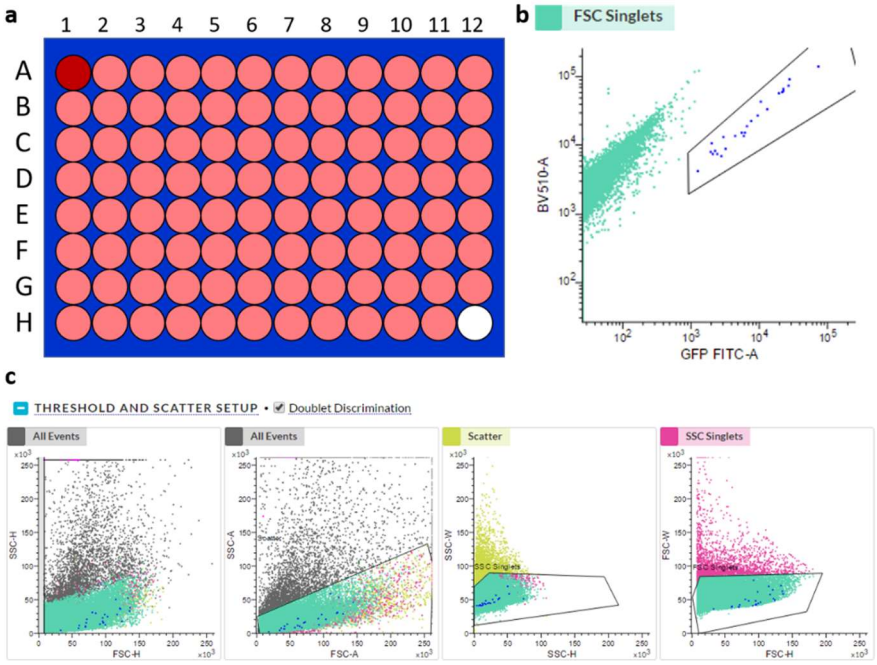


Figure 3-2: Sorting strategy using a BD FACS-Melody. **A**) Plate setup for sorting. Red wells (A1) contained 20 cells as a positive control, pink wells contain single cells and white wells (H12) were deliberately left blank. **B)** Tapetal cells were sorted by selecting cells with high GFP fluorescence (x-axis) relative to an absent fluorophore (y-axis). Non-GFP positive cells cluster along a straight line when GFP fluorescence is plotted against an absent fluorophore (background fluorescence). GFP-positive cells are shifted right on the plot. **C)** Doublet discrimination was employed to ensure only individual cells within one droplet were sorted. Doublets of cells have a higher width (y-axis) relative to height (x-axis) than singlets.

A total of 288 libraries, 282 of which were single-cell, were sequenced on an Illumina HiSeq 2500 (Appendix table A1). This was to avoid index switching (the incorrect assigning of sequencing reads) which occurs with the ExAmp chemistry of Illumina HiSeq 3000/4000/X-ten models [269]. Libraries were sequenced to a mean depth of

682,260 reads, but these were unevenly distributed across samples. The number of genes detected correlated positively with sequencing depth, with an average of 3892 genes detected per cell (Fig. 3-3). The distributions of both metrics are positively skewed as there were many cells with very few reads (and hence few genes detected) (Fig. 3-3). For further analyses, cells with more than 250,000 read counts and 1,500 genes expressed were selected, giving 141 single cells passing quality control.

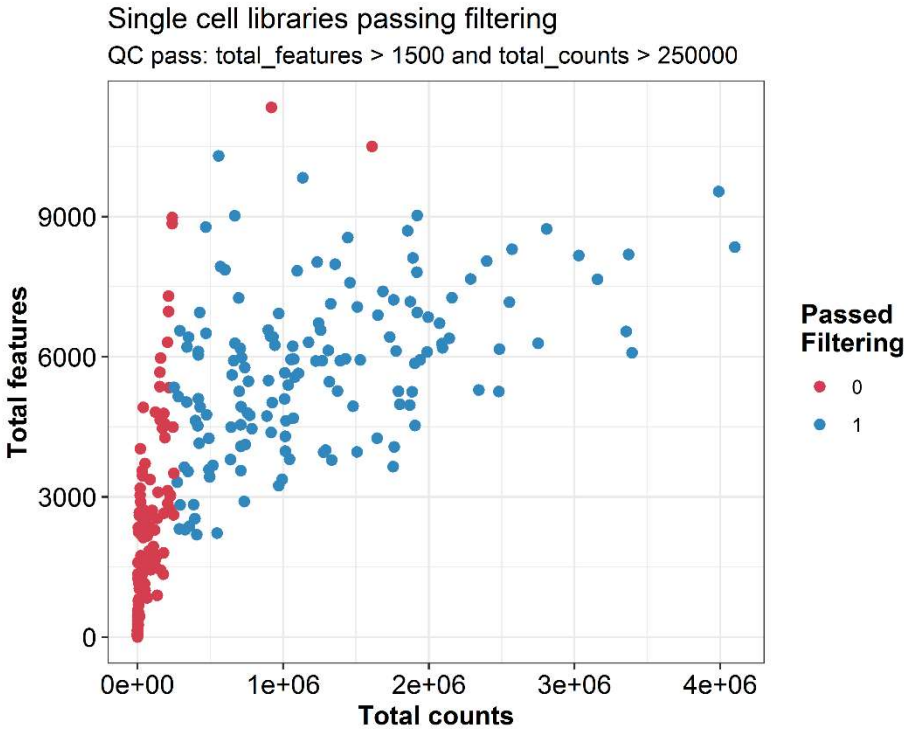


Figure 3-3: Read count and the total number of genes detected for all sequenced libraries. Libraries made from single cells with more than 250,000 reads and 1,500 genes detected (blue) were used for analysis in Monocle and the manual approach (SC3).

To analyse the quality of single-cell mRNA-seq libraries, gene expression in single-cells was compared with bulk tapetal mRNA-seq data collected previously (Chapter 2, Appendix table A1). The 100 most highly expressed genes from both bulk and single-cell mRNA-seq datasets are similar, with 63% shared between them. These lists include genes functioning in sporopollenin and lipid metabolism, indicating that, as in Chapter 2, high purity tapetal cells have been collected. Global gene expression shows very high correlation between bulk sequencing and single-cell libraries (R^2 = 0.959, Pearson's product-moment correlation) (Fig. 3-4), suggesting highly similar

populations of cells were sorted for each set of experiments. This demonstrates that combined single-cell mRNA-seq libraries are a good approximation to bulk mRNA-seq data, but with the further ability to examine heterogeneity in gene expression.

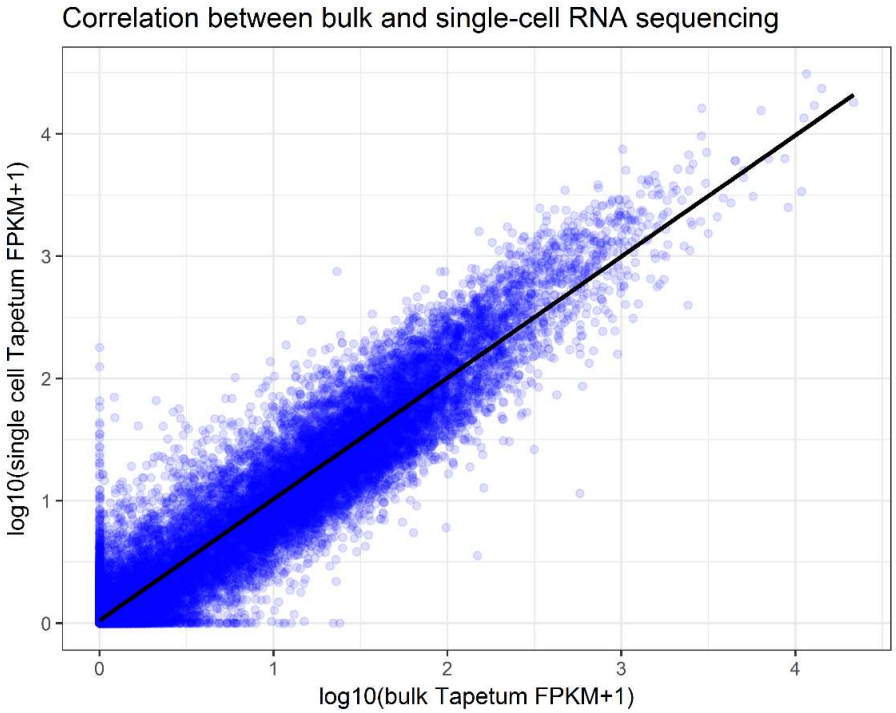


Figure 3-4: Correlation between bulk and single-cell RNA-sequencing batches. Three independent biological replicates for bulk RNA-seq were compared with three plates of single-cell libraries. $R^2 = 0.959$, Pearson's product-moment correlation.

3.3 Pseudotime ordering can be used to infer changes in gene expression over tapetal development

Sequencing individual cells from a developmental series allows differences between different staged cells to be visualised, rather than an average of many distinct stages. Collecting individual cells from a tissue necessarily destroys positional and developmental information, meaning this must be inferred from measurements of the transcriptome. However, recreating developmental trajectories and identifying different cell states is not a trivial problem, and many different approaches are employed to tackle this issue. I have therefore employed three different strategies to recreate tapetal gene expression over developmental time, two R packages, Monocle

[267] and SCORPIUS [270], as well as a manual principal component analysis-based approach.

3.3.1 Monocle fails to reliably recreate tapetal gene expression through pseudotime

Monocle is one of the oldest single-cell pseudotime packages, and as such is a popular choice of software for analysis. To order cells in pseudotime, Monocle reduces the dimensionality of the data through Independent Component Analysis, and then plots a route through the cells via a Minimum Spanning Tree (MST). An MST is the subset of edges (pseudotime steps) which connects all vertices (cells), with no cycles and the minimum total edge weight (step distance) [271]. Creating an MST for tapetal single-cell data yields the following representation of cells divided into three states (Fig. 3-5).

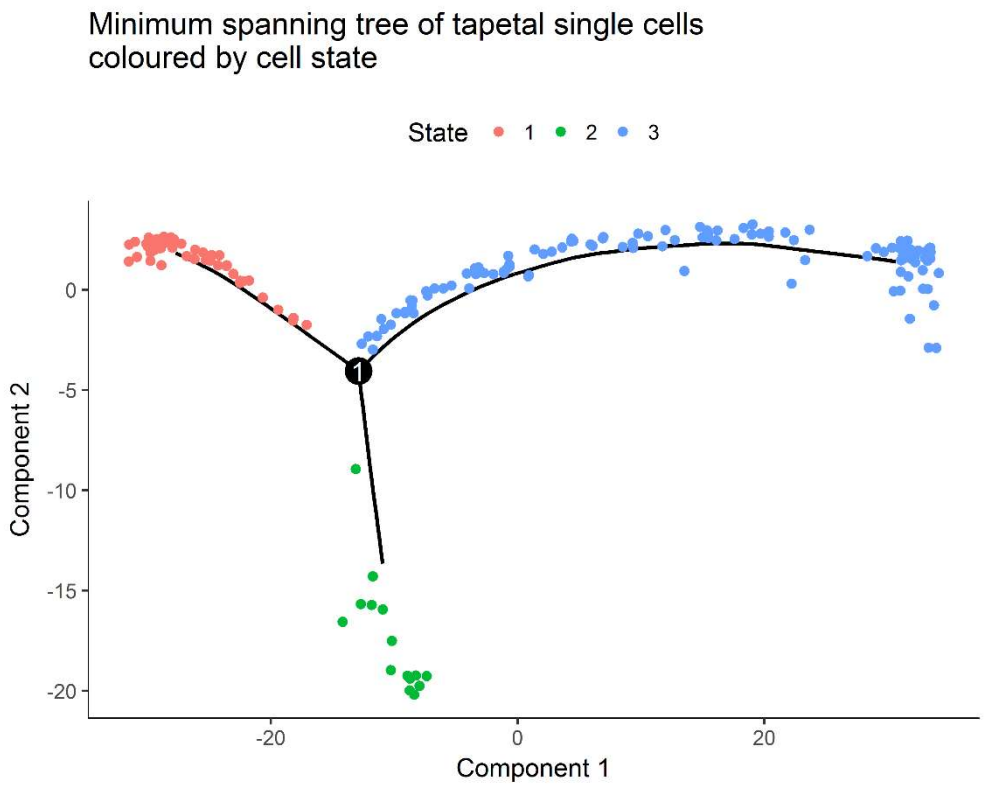


Figure 3-5: MST output from Monocle. Tapetal cells are divided into three cell states. State One represents the start of pseudotime with states Two and Three as alternative end states. Pseudotime is represented by Component one. How two alternative ends states can be reconciled with tapetum biology is unclear.

The progression of pseudotime shows that cells in state One are the earliest, then state Two, and state Three are at a developmentally later stage (Fig. 3-5). This clustering of cells is difficult to reconcile with what is known about tapetum development. The tapetum is a tissue which shows coordination between cells at the same stage and eventually enters programmed cell death (PCD) [10]. How a rare alternative cell state could arise when all cells will progress to PCD is hard to envisage. State Two could represent a pause state through development, though evidence for this is lacking [14].

Expression of marker genes shows that Monocle is not able to recreate known expression profiles. *SPOROCYTELESS/NOZZLE* (*SPL/NZZ*) functions in early anther development and becomes progressively restricted to the germ cells over time [272]. *SPL* is no longer expressed in the tapetum after anther stage 8, but *SPL* shows no variation along the Monocle pseudotime axis (Fig. 3-6). The polyketide synthase gene *LESS ADHERENT POLLEN 6* (*LAP6*) functions in sporopollenin production and is only expressed late in tapetal development from stage 8 [49]. *LAP6* does show a late-stage expression profile overall, but many cells later in pseudotime show no expression while earlier cells do (Fig. 3-6). This differs from published data, which shows all cells express *LAP6* at high levels in late stage tapetum [49].

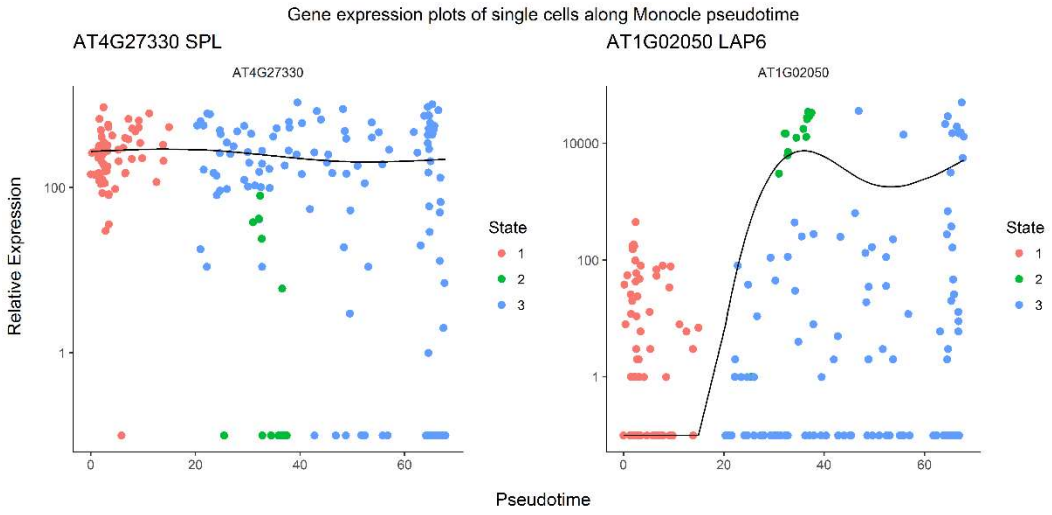


Figure 3-6: Expression of *SPL* (*AT4G27330*) and *LAP6* (*AT1G02050*) through Pseudotime. Neither expression profile matches published data. In vivo *SPL* declines over tapetal development and *LAP6* is expressed late in tapetal development and is highly expressed in all late stage cells.

Investigating pseudotime variation with cell sequencing metrics showed that pseudotime negatively correlates with both read count and total number of genes detected (Appendix Fig. A2). Cells in state One are sequenced to a greater depth and have more genes detected per cell, while cells in state three have the fewest reads and genes expressed. This occurs after processing by Monocle, even though it is suggested data need not be normalised before analysis. Monocle is therefore forming spurious correlations between biological and technical signals. Data must be normalised to remove variation with sequencing depth and yield usable results from pseudotime analyses.

3.3.2 A manual approach can be used to yield an estimate of developmental time

All pseudotime analyses rely on two processes to extract an estimation of developmental time:

1. Reducing the dimensionality of the data through one of several dimensional reduction techniques, and
2. Finding relationships between cells in this reduced dimension space that may represent a biological process.

Both steps are possible to implement without the need for specialist software. To avoid spurious correlations between cell sequencing quality and pseudotime, the data was first normalised. The log expression, in transcripts per million (TPM), of each gene was correlated against the total number of genes expressed per cell and the residuals of expression taken.

Normalised expression data was then analysed using the SC3 R package [273]. After normalisation, the percentage variance explained by the first principal component is far lower than for raw expression data (PC1 76% variance to 10%). This would suggest that the majority of variance in the transcriptomes was due to differences in sequencing depth and the number of genes detected between cells. As was the case

with Monocle, any attempt to extract a biological signature from raw data will be confounded by technical noise.

From silhouette widths [274], the data was divided into two clusters separated along PC2 (Fig. 3-7). SC3 was used to extract marker genes for each cluster at $p < 0.05$. Cluster one had 34 associated marker genes, and the second cluster had 69. No significantly enriched Gene Ontology (GO) terms were associated with marker genes from cluster one, but cluster two was enriched for terms such as "pollen exine formation" (GO:0010584, $p = 1.9 \times 10^{-8}$) and "pollen development" (GO:0009555, $p = 1.78 \times 10^{-5}$) [189-191]. This suggests that cluster two contains late-stage tapetum, expressing sporopollenin-biosynthesis genes, and that principal component two represents an approximation to a developmental time axis (Fig. 3-3).

Given this division of cells, and the late tapetal marker expression in cluster two, loadings of principal component one and two were extracted to plot cells in order along the axes. Raw expression data (TPM) was used to plot gene expression profiles (Fig. 3-8). *SPL* shows an early expression profile along PC2 while *LAP6* is nearly undetectable early along PC2 but rapidly increases in expression. This suggests that this manual approach is able to recreate expression of early- and late-stage genes in the tapetum.

While this manual approach has worked well at distinguishing expression profiles of broadly early and late genes, specifically those involved in pollen exine formation, there is little distinction between cells at the stages prior to sporopollenin production. If the best approximation of the pseudotime axis is not parallel to PC1 or PC2, then this manual approach will only produce a rough estimate of gene expression changes. Ideally an approach would be able to ensure spurious correlations with technical noise are not formed, but the greatest degree of variation is captured by the pseudotime axis inferred.

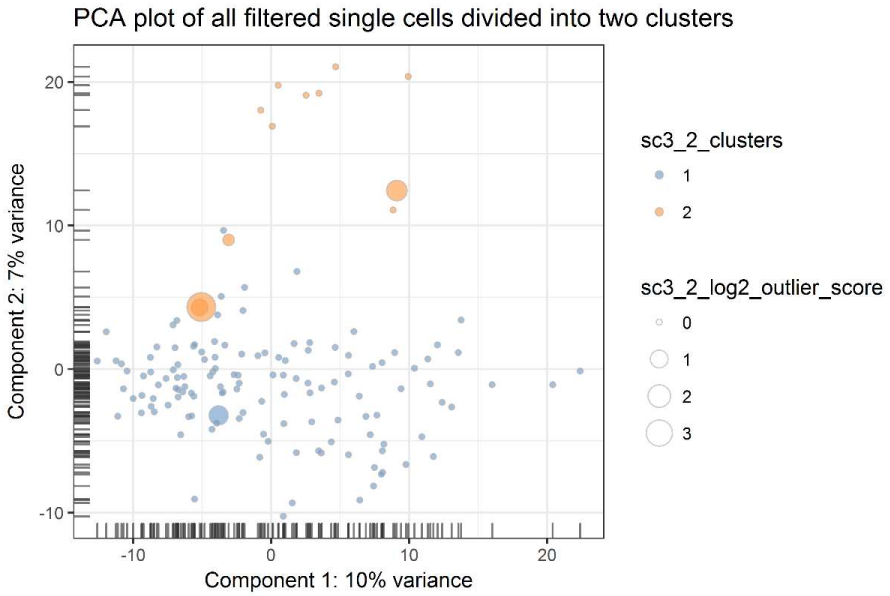


Figure 3-7: PCA plot of all single-cell libraries passing filtering. Cells are divided into two clusters along principal component two. Late-stage marker genes were found to be highly expressed in cluster two and lowly expressed in cluster one, suggesting that component 2 represents pseudotime. Point size represents the outlier score for cells in each cluster.

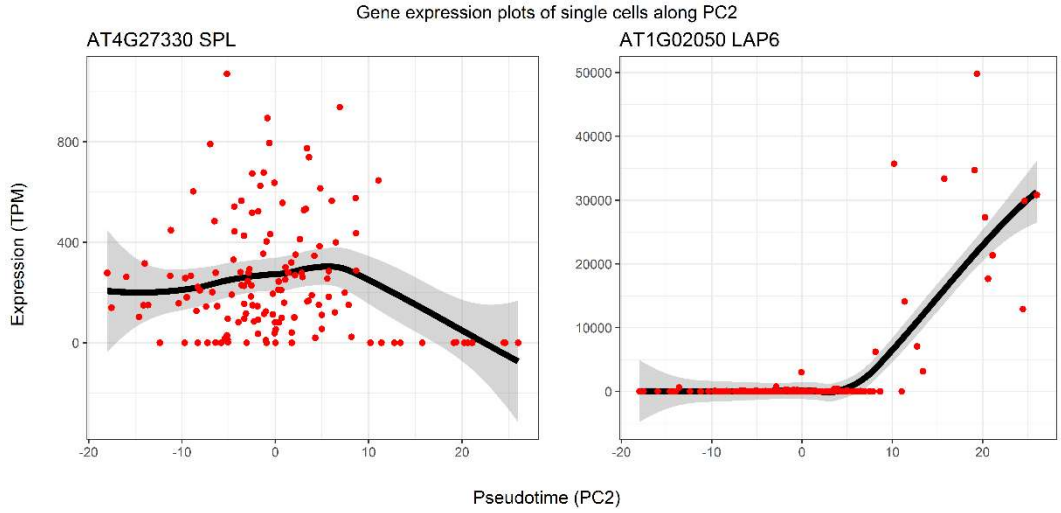


Figure 3-8: Expression of two marker genes along principal component two. *SPL* shows a noisy expression profile but declines in expression through tapetal development as expected. *LAP6* shows a sharp induction only at late stages. Points represent expression in individual cells and black lines represent LOESS fitted curves, bounded by a 95% confidence interval.

3.3.3 SCORPIUS produces a reliable pseudotime axis which matches known tapetal gene expression

SCORPIUS is a pseudotime trajectory inference package which performs well against others in reliably reproducing cell orders [270]. SCORPIUS requires no *a priori*

information to order cells. Correlation distances are calculated between all cells and outliers are removed to reduce their negative impact on the inferred trajectory. Dimensionality is reduced through multi-dimensional scaling. An initial trajectory is calculated by finding the shortest path through clusters of cells, grouped by k-means clustering. This trajectory is then refined iteratively using the principal curves algorithm [275]. SCORPIUS then infers the extent to which a gene is involved in the process of interest (e.g. cell differentiation). This is achieved through ranking all genes according to their ability to predict cell order from the expression data, by implementing the Random forest algorithm [276]. This allows genes to be clustered into modules of gene expression [270].

An advantage of SCORPIUS is that branch points are not forced into trajectories. This works well for tapetal single-cell data, as it is known that there are no alternative end cell states in tapetum development. Another advantage is the user control over trajectory inference. The number of genes used to infer the trajectory can be altered and the best fitting trajectory selected. For tapetal data, I found that the trajectory inferred from all filtered cells (183) was more reliable than that created by selecting the most variable genes (data not shown). The trajectory shown in the rest of this chapter is that inferred from the dimensionally-reduced filtered expression data. (Fig. 3-9)

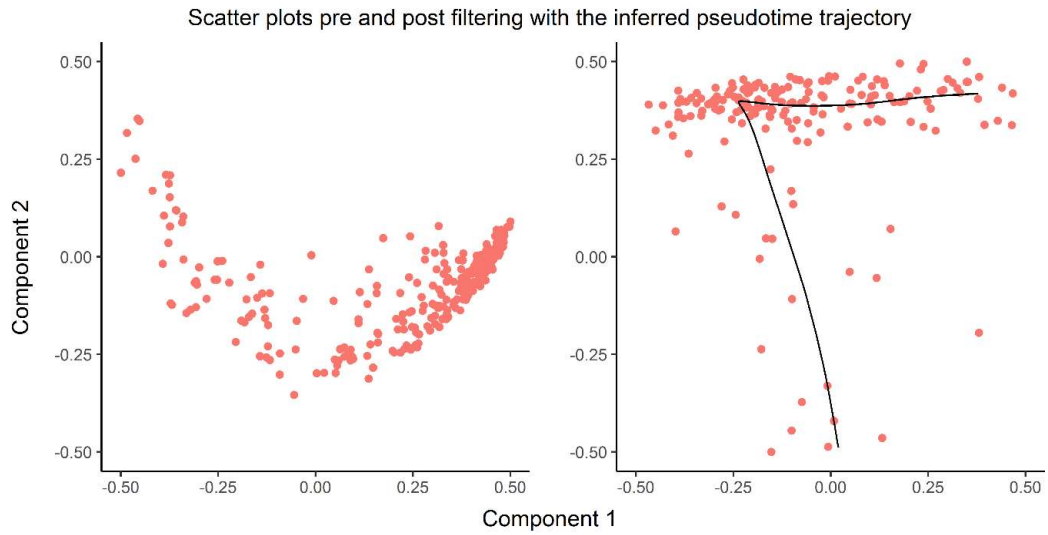


Figure 3-9: Dimensionally reduced representation of tapetal single-cell libraries pre- and post-filtering by SCORPIUS. The black line represents the trajectory inferred from all filtered cells, starting at the top right ($t = 0$) and ending at the bottom of the plot ($t = 1$).

To validate the cell trajectory inferred, the expression of genes expressed at known tapetal stages were investigated. Selected marker genes, while variable, do corroborate the inferred cell trajectory (Fig. 3-10). *SPL* shows a matching early expression profile in pseudotime. The MYB transcription factor gene *DEFECTIVE in TAPETAL DEVELOPMENT and FUNCTION1* (*TDF1*) is expressed in the tapetum from anther stage 5 to 7 [83, 84]. In pseudotime *TDF1* shows a noisy expression profile, but also appears to decline in expression at a similar time point to *SPL* (Fig. 3-10). The *bHLH010* gene encodes a *DYSFUNCTIONAL TAPETUM* (*DYT1*)-interacting transcription factor that is strongly expressed at stages 6 and 7, but is unexpressed from stage 9 [82]. In pseudotime *bHLH010* shows two peaks of expression and a decline after $t \approx 0.6$ (Fig. 3-10).

The clearest genes are those that are highly expressed in late stages such as genes involved in sporopollenin production. *LAP6* [49], *MALE STERILITY 2* (*MS2*) [277], and the cytochrome P450 *CYP703A2* [51] all function in sporopollenin biosynthesis and are expressed from anther stage 8 after tetrad release. In pseudotime, all three genes are undetected, until they are induced at $t \approx 0.6$ (Fig. 3-10).

The good correspondence between the decline in early expressed genes and the induction of late genes suggests that the pseudotime axis is oriented correctly and

that $t = 0.6$ reflects the start of anther stage 8 [10, 11]. While differences between early and late tapetum expression can be visualised, this data may not be able to distinguish subtle differences between tapetum stages, particularly before the tapetum begins to produce pollen wall material (Stage 8).

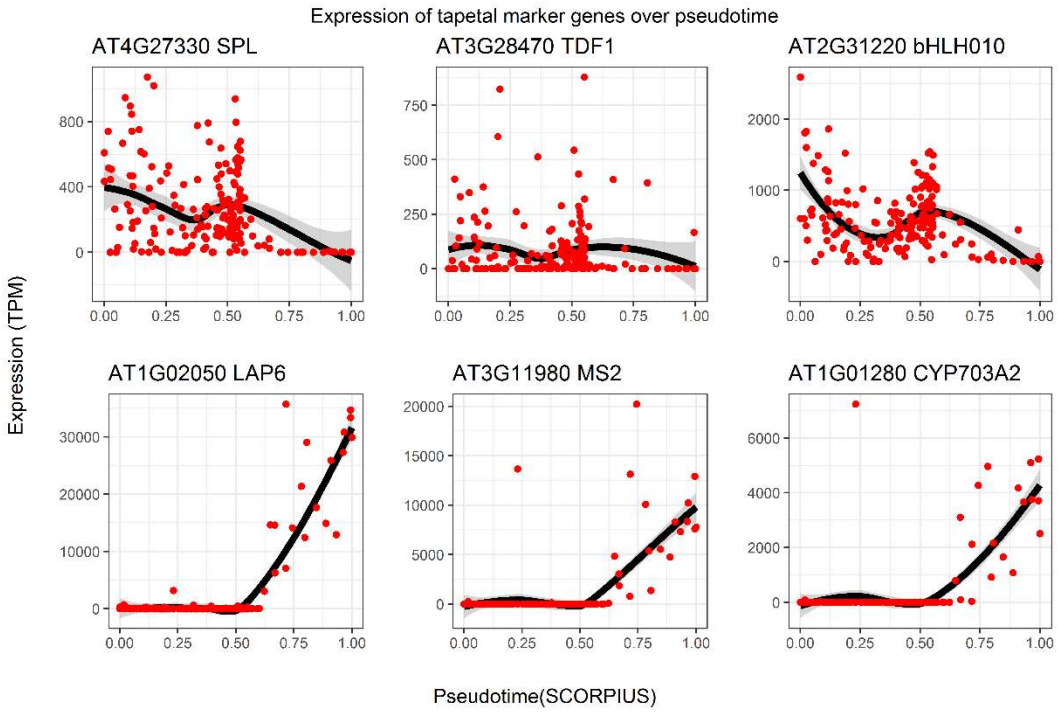


Figure 3-10: Expression of six tapetal marker genes. The early expressed genes *SPL*, *TDF1*, and *bHLH010* show a decline in expression over pseudotime. *bHLH010* may show a biphasic expression which has not been reported previously. Late tapetal genes functioning in sporopollenin biosynthesis; *LAP6*, *MS2*, and *CYP703A2* show expression only late in tapetal pseudotime. Red points represent expression in individual cells and black lines represent LOESS fitted curves bounded by 95% confidences intervals.

3.4 Variably expressed genes can be divided into two expression modules

SCORPIUS can divide genes, which show variable expression, into groups based on their expression profiles and levels [270]. From the 1000 most variably expressed genes, SCORPIUS produced four modules of gene expression. Modules two, three, and four showed the same expression profile however, and only differed in their expression level. They were then combined to produce a new Module two. This gave two modules with 758 genes in Module one and 242 genes Module two. The mean expression of all genes in Module one shows a decline across tapetal pseudotime,

without reaching zero (Fig. 3-11). The absolute mean change in expression, however, is subtle. Mean Module two expression shows a much more dramatic change in expression, with nearly all Module two genes being undetected early in pseudotime and reaching much higher levels than Module one genes later in tapetal pseudotime (after $t = 0.6$) (Figs. 3-10 & 3-11).

Significantly enriched Gene Ontology (GO) terms for each cluster, reveal that Module two genes are predominantly involved in pollen wall formation. Module two is enriched for GO terms such as "pollen exine formation" (GO:0010584, $p = 2.44 \times 10^{-11}$), and "fatty acid biosynthetic process" (GO:0006631, $p = 8.51 \times 10^{-5}$)[189, 191]. Module two contains known sporopollenin biosynthesis genes such as *CYP703A2*, *LAP5&6*, and *ABCG26*, as well as fatty acid biosynthesis genes such as *MS2*, *ACYL-COA SYNTHETASE 5* (*ACOS5*), and *FATTY ACID DESATURASE 5* (*FAD5*).

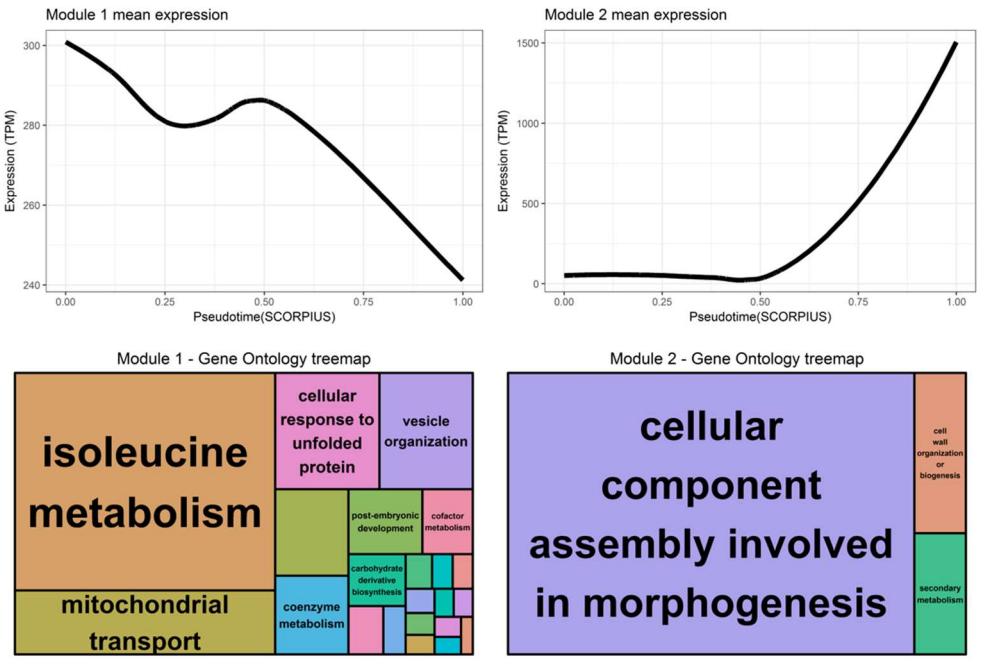


Figure 3-11: Mean expression profiles along pseudotime for genes in Module one and two. Lines are averages of all genes found in each expression module. Module one shows a decline in expression over pseudotime, though the average difference in expression is relatively small. Module two shows no expression early in pseudotime and are only expressed from $t = 0.6$, often to a high level. Treemaps of significantly enriched GO terms for each expression module, show more GO terms associated with Module one, though they are less significant than those associated with Module two. Lines represent LOESS fitted curves.

More GO terms are associated with Module one, reflecting the larger number of genes expressed though they are less significantly enriched than terms associated with Module two. GO terms associated with membrane trafficking, such as "ER to Golgi vesicle-mediated transport" (GO:0006888, $p = 2.36 \times 10^{-8}$) are enriched in Module one, reflecting the tapetum's role as an excretory tissue building and degrading cell walls. GO enrichment analysis also suggests a wave of epigenomic reorganisation during early tapetum development as GO terms associated with "chromatin assembly or disassembly" (GO:0006333, $p = 0.000485$), "chromatin organization" (GO:0006325, $p = 0.00181$), and "gene silencing by RNA" (GO:0031047, $p = 0.0006$) are significantly enriched. Several shared Pol IV/V components; *NRPB5*/*NRPD5*, *NRPD2A*/*NRPE2*, *NRPB8B*/*NRPD8B*/*NRPE8B*, and *NRPB6A*/*NRPD6A*/*NRPE6A* are also present in Module one, possibly suggesting a role for RNA-directed DNA methylation at early stages of tapetum development.

There is a high degree of overlap between SCPORPIUS and the manual approach for late pseudotime expressed genes. Of the genes identified as late pseudotime expressed in the manual approach, 59 of 68 (87%) were also identified in Module two by SCORPIUS. SCORPIUS was able to identify many more genes however (240). The number of genes identified as early pseudotime expressed varies greatly, with 747 identified in SCORPIUS and 34 through the manual approach. Only five genes overlap between the lists, suggesting that the definition of early pseudotime varies most between methods.

Differences in expression can be used to link pseudotime to defined anther stages more confidently than the expression of a single marker gene [10]. Having two expression modules with decreasing or increasing expression over pseudotime suggests it can be divided into two stages, early and late (Figs. 3-10 & 3-11). I can therefore be confident in defining $t = 0.6$ as the onset of anther stage 8. Later stages are harder to identify, due to the small number of cells later in pseudotime and the lack of expression of programmed cell death markers (e.g. *CEP1*, data not shown) [18, 73]. As genes such as *ABORTED MICROSPORES* (*AMS*) and *MALE STERILITY 188*

(*MS188*) are still expressed late in pseudotime, when they are absent in stage 10 tapetum (data not shown) [80], suggest that the latest cells sequenced are stage 9. Late pseudotime (t = 0.6-1) can therefore be define as stages 8-9.

Early pseudotime is therefore pre-stage eight, but at what stage the earliest cells collected are is hard to decipher. Tapetal cells are defined at stage five, but may be harder to collect at this stage due to the smaller size of the anthers [10]. The noisy expression profile of early marker genes such as *TDF1* also make this challenging (Fig. 3-11). The lack of expression of *DYT1* (data not shown), but expression of downstream genes such as *TDF1* and *AMS* (Fig. 3-11) suggests that the earliest cells are from stage six [80, 84]. It is possible, however, that *DYT1* may be too lowly expressed to detect even in cells of stage 5. Given the lack of evidence of earlier stage cells it is most likely that early pseudotime (t = 0-0.6) represents anther stages 6-8. Given the differences in module expression profile, it appears that the onset of stage 8 represents a developmental switch in terms of tapetum transcription.

This inference is limited, however, as stages are defined by morphological features, and gene expression has largely been defined through *in situ* hybridisation. scRNA-seq may be more sensitive than *in situ* hybridisation to detect variation in expression and so reveal gene expression dynamics which do not necessarily correlate with defined anther stages.

3.5 The number of genes expressed varies with pseudotime

As well as testing gene expression profiles of individual genes, I can also use this pseudotime order of cells to investigate other changes over tapetum development. From sequencing, I also have information on the number of genes detected in each cell. The number of genes expressed per cell shows variation over pseudotime (Fig. 3-12). In early pseudotime (t = 0-0.25) the number of genes expressed increases, from an average of roughly 3000. At t = 0.325 there is a step change to a much higher number of genes expressed. After t = 0.6 the total declines as cells enter late tapetum

development and expression of sporopollenin biosynthesis genes increases. This would suggest that tapetal cells express the most genes just prior to stage eight and express the fewest at stage 6.

Whether this phenomenon is truly biological or if this is an artefact of single-cell sequencing and analysis is unknown. The number of genes expressed does correlate with read count early in pseudotime, though this relationship is reversed later (Appendix Fig. A3). It is possible that cells progress through a phase of enhanced gene expression as they differentiate, before declining as the cells complete their function and begin expressing PCD-related genes. Changes in total gene expression have been detected in other systems. For example, it is known that changes in the number of genes expressed accompanies development in zebrafish blood cell differentiation [264]. This suggests that changes in total gene expression may commonly occur in developmental processes.

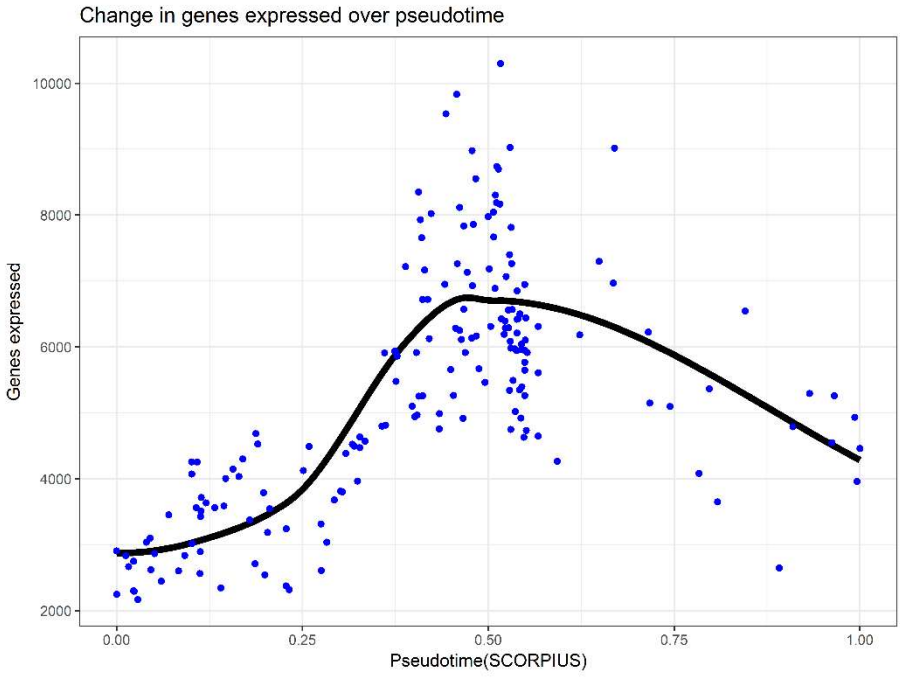


Figure 3-12: Changes in the number of genes expressed through pseudotime. The number of genes expressed appears to depend on the stage of the cell. The number of genes expressed shows three phases through pseudotime, a step change from low to high number of genes expressed occurs at around $t = 0.3$, and then slowly declines as the tapetum ages from around $t = 0.6$. Blue points represent the number of genes expressed in individual cells and the black line represents a LOESS fitted curve.

3.6 Gene expression modules highlight novel stage expression of genes

As well as genes known to play a role in pollen development, pseudotime ordering can reveal novel genes which may also function in tapetal development. Module one contains several exciting groups of genes which have not previously been identified to function early in tapetum development. Several DNA topoisomerase genes are expressed early in tapetum development where they may function in cell division and endoreplication [35]. As well as *DNA TOPOISOMERASE 1 ALPHA* (*TOP1α*) and *TOP1β*, *TOP6B* is expressed in mid pseudotime (Fig. 3-13). *TOP6B* has been shown to be essential for endoreplication [278], and expression in mid-pseudotime suggests that endoreplication occurs in these cells. The Cyclin-dependent kinase inhibitor, *INHIBITOR OF CDC2 KINASE 4/KIP-RELATED PROTEIN 6 (ICK4/KRP6)*, is known to block cell cycle progression and is expressed late in tapetal development (Module two) [279]. This may function to suppress tapetal cell division or may have a role in the tapetum outside of cell division control. Together this suggests that there is tight regulation of endoreplication and cell cycle progression in the tapetum, which can be investigated through single-cell sequencing (Fig. 3-13).

Two interesting candidates from Module two are the small peptides *RAPID ALKALISATION FACTOR LIKE1* (*RALFL1*) and *CLAVATA3/ESR-RELATED 1* (*CLE1*) (Fig. 3-13). Small peptide signalling in early anthers controls the positioning and proliferation of anther cells types [14, 34]. However, such signalling at later anther stages has not been described. The late stage expression of *CLE1* and *RALFL1* (post stage 8) suggests that the tapetum signals to post-meiotic sexual lineage cells (microspores and pollen) through small peptides. Studying these peptides in developing anthers may reveal new signalling pathways controlling male germline development.

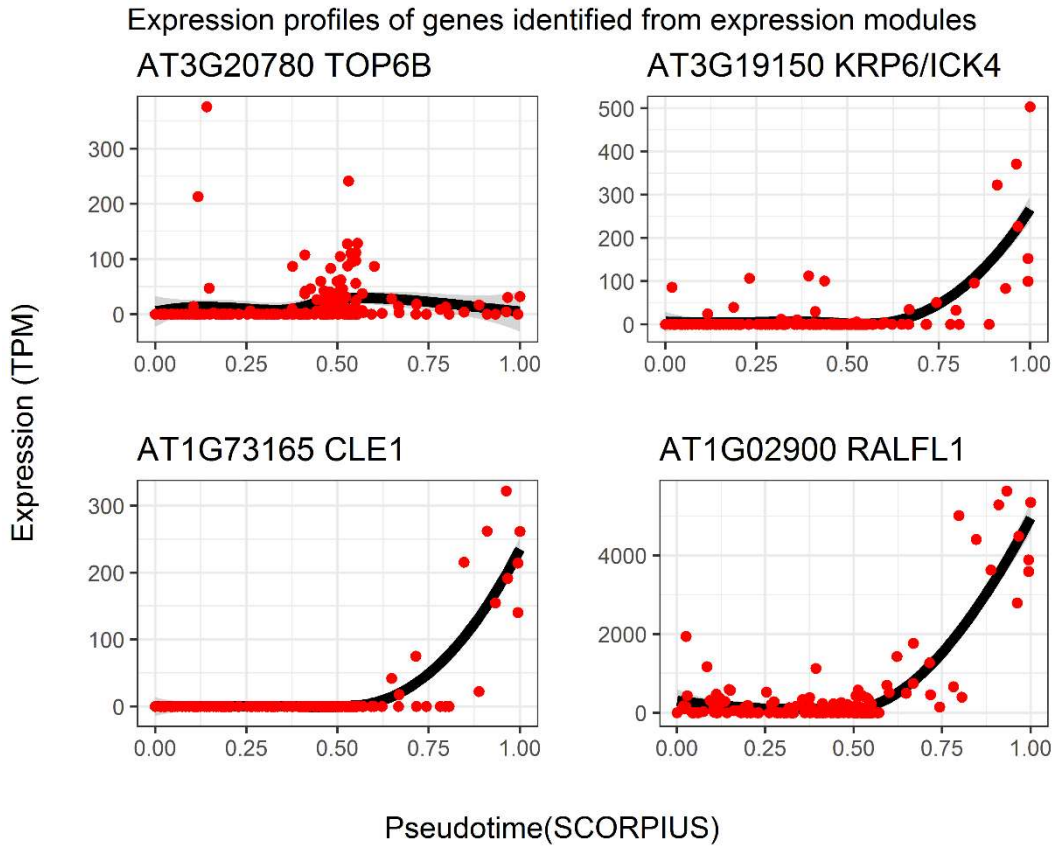


Figure 3-13: TOP) Novel gene expression profiles of endoreplication-related genes *TOP6B* and *KRP6/ICK4*. *TOP6B* is classified in Module one and *KRP6/ICK4* in Module two. The timing of *TOP6B* expression is suggestive of endoreplication, whereas *KRP6/ICK4* is known to inhibit cell cycle progression. **BOTTOM**) Expression profiles of two small signalling peptides, *CLE1* and *RALFL1* from expression Module two. Both small peptide genes show a late-stage expression profile, suggesting they function in signalling after stage 8 ($t = 0.6$). Red points represent expression in individual cells and black lines represent LOESS fitted curves.

3.7 Tapetum-enriched and tapetum-specific transcription factors show stage-specific expression profiles

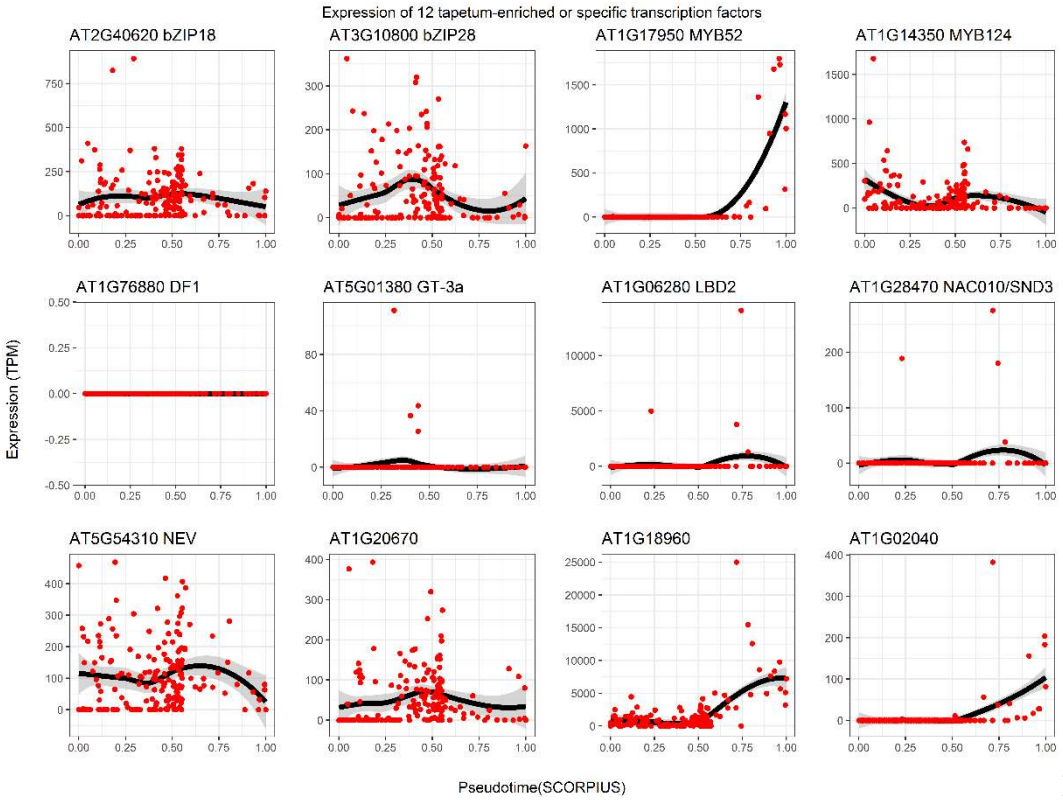
In the previous chapter I identified tapetum-enriched transcription factors and tapetum-specific genes from bulk RNA-sequencing data (Chapter 2). Now with ordered single-cell transcriptomes I investigated the developmental timing of expression (Fig. 3-14). This can be correlated with phenotype to aid the prediction of downstream target genes.

Some of these transcription factor genes have already been identified as variably expressed through tapetal pseudotime. Three have been classified into expression Module two; *MYB52*, *AT1G02040* encoding a C2H2 Zn-finger protein, and the MYB-like helix-turn-helix protein gene *AT1G18960*. *AT1G02040* has previously been identified as a downstream target of MS188, which is corroborated by pseudotime data (post stage 8)(Fig. 3-14) [204]. *MYB52* has been shown to regulate pectin demethylesterification in seed mucilage [199], which is also an important process for microspore release [42]. This may therefore suggest a role for MYB52 in regulating pectin catabolism in pollen and may warrant a closer inspection of *myb52* pollen phenotype.

LBD2 is undetected in nearly all cells but is very highly expressed in a few late-stage cells, supporting proposed regulation by MS188 (Fig. 3-14) [204]. *LBD2* expression appears markedly similar to the profile of *NAC DOMAIN CONTAINING 10/ SECONDARY WALL-ASSOCIATED NAC DOMAIN PROTEIN 3* (*NAC10/SND3*), and when the two are compared in individual cells, are strongly correlated ($R^2 = 0.77$, Pearson's product-moment correlation) (Fig. 3-14). This supports the predicted role of LBD2 in regulating expression of *NAC10/SND3* from DAP-seq binding data [218]. The high level of expression of these genes in few cells suggests rapid transcription and mRNA turn-over. These data fit with the observed tapetal phenotype of *lbd2*. Highly vacuolated tapetal cells are seen from stage 9, while the tapetum appears to be WT at earlier stages (Chapter 2). The co-expression of *LBD2* and *NAC10/SND3* supports a proposed role of *LBD2* in regulating pollen wall biosynthesis (Chapter2) as *NAC10/SND3* has been shown to function in secondary cell wall biosynthesis [205, 206]. While a *nac10/snd3* mutant did not show the tapetal phenotype observed in *lbd2*, pollen phenotype was not assayed. Regulation of *NAC10/SND3* by another factor may explain the variation of pollen phenotype seen in the *lbd2* mutant.

The bZIP transcription factors *bZIP18* and *bZIP28* show little variation in expression through pseudotime (Fig. 3-14). While the *df1* mutant shows both tapetal and pollen defects, expression in wild-type single-cells was not detected (Fig. 3-14). This may

suggest that *DF1* expression occurs very briefly through tapetal development, and *DF1*-expressing cells have not been collected, or that *DF1* is expressed at an earlier or later stage than those represented by pseudotime.



F

Figure 3-14: Expression of 11 tapetum enriched and one tapetum-specific (*LBD2*) transcription factors identified from bulk RNA-seq. Pseudotime expression profiles and co-expression can suggest possible interacting partners and downstream targets. *MYB52*, *AT1G02040*, and *AT1G18960* are classified in Module two. Red points represent expression in individual cells and black lines represent LOESS fitted curves.

The expression of other tapetum-specific genes (Chapter 2) was investigated to find genes which have stage-specific expression. Late-stage specific genes have been identified (e.g. *CYP703A2* Fig. 3-10), but strong and early tapetum-specific genes have been harder to identify. The cytoplasmic tRNA 2-thiolation protein *AT3G55570* is strongly expressed in early stage tapetum and is undetected after inferred stage 8 (Fig. 3-15). The genes *AT1G23670* and *AT3G26140* show expression only prior to inferred stage 8 (Fig. 3-15). *AT4G22870* is a late-stage tapetal gene but shows a later increase in expression than sporopollenin-biosynthesis genes (Fig. 3-15). While the functions of these genes may not be understood, their tapetum- and stage-specific expression could prove to be a powerful tool in creating transgenic plants with stage-

specific reporter, overexpression, or complementation constructs. This will allow us to further dissect specific tapetal functions in *Arabidopsis*.

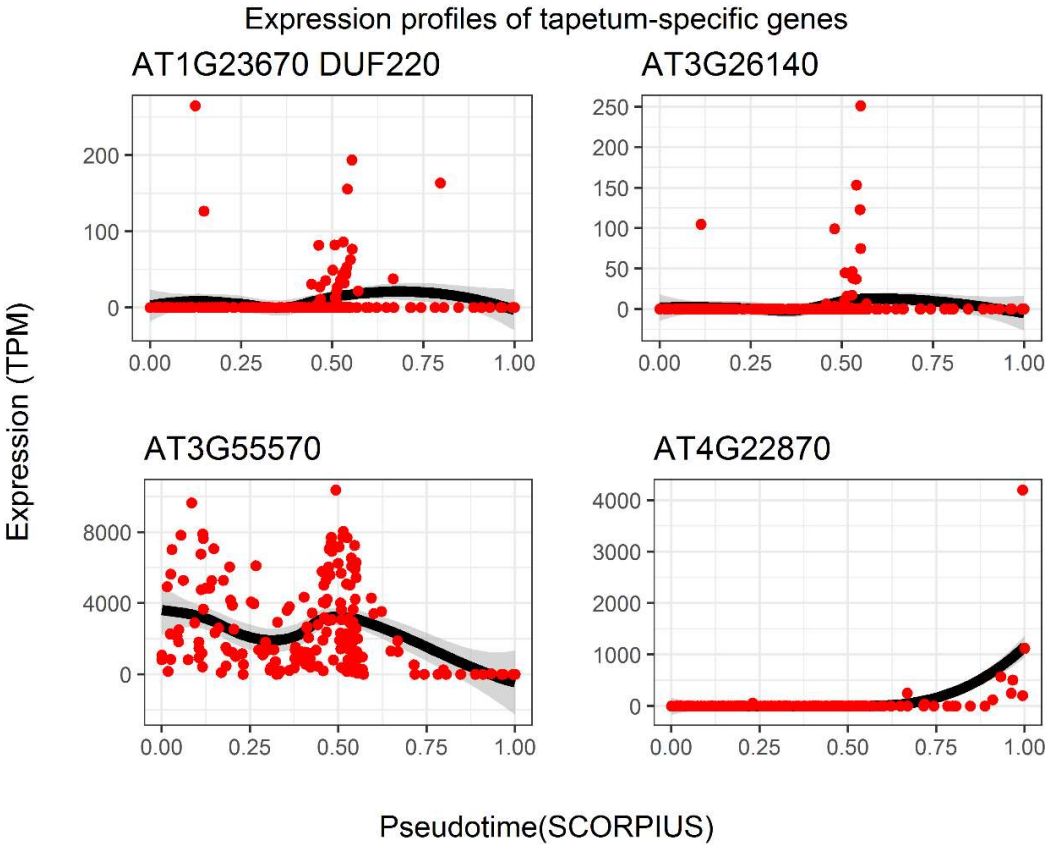


Figure 3-15: Expression of four tapetum-specific genes which also show stage-specific expression. *AT1G23670* and *AT3G26140* show mid-pseudotime expression, before the start of inferred stage 8. *AT3G55570* is expressed strongly only in pre-stage 8 cells and declines late in pseudotime. *AT4G22870* shows a late induction of expression, though after genes such as *LAP6*, suggesting this is a later stage marker. Red points represent expression in individual cells and black lines represent LOESS fitted curves.

3.8 Discussion

Understanding gene expression at the level of single cells can reveal novel sub-populations, gene expression dynamics and regulatory interactions. Applying single-cell mRNA sequencing to FACS-sorted tapetal cells has produced gene expression data which compares well with bulk mRNA-seq. It has allowed the identification of stage-specific expression of tapetal genes, as well as genes not previously identified functioning in the tapetum. These results have led to exciting hypotheses about tapetum function which require further testing.

3.8.1 Single cell sequencing can be used to recreate developmental changes in gene expression

While many software packages are available to analyse single-cell data, and specifically pseudotime data, care must be taken in selecting packages. Both the manual approach and SCORPIUS, produced pseudotime axes that match the known expression profiles of tapetal genes. While the manual approach worked well for finding the expression of late-stage genes, SCORPIUS was able to find many genes which are expressed both early and late in tapetal development. This has shown novel expression profiles of both known tapetal genes, and those that have not been previously implicated in tapetum biology.

A developmental switch was inferred at the onset of anther stage 8, with large-scale changes in gene expression in the tapetum. Pseudotemporal ordering of tapetal cells also suggests that the tapetum progresses through distinct phases of gene expression, with the number of genes expressed changing through tapetal development. It is possible that this is an artefact of sequencing variation and clustering but gives an exciting hypothesis to test. While bulk sequencing of mixed-stage tapetum would never be able to identify these changes, sorting cells of specific stages, with the markers identified (section 3.7), may be able to test this phenomenon.

3.8.2 Pseudotime expression profiles provide candidates for novel tapetal genes and functions

Single-cell analysis was used to infer stage-specific expression of genes, without the need to collect cells from manually staged flowers. A temporal understanding of gene expression is a valuable resource when investigating gene function. Many of the transcription factors identified as tapetum-enriched or tapetum-specific (Chapter 2) show stage-specific expression. This information, combined with published data, can help elucidate the role of these transcription factors in the tapetum. The expression of *LBD2* corroborates both published expression data, and the phenotypes observed in the tapetum and pollen, corroborating a role in sporopollenin production (Chapter

2). Whether the variable expression of *LBD2* seen at the single cell level occurs within cells of the same locule is unknown. Stochastic expression may partly explain the rare severe defects in *lbd2* anthers (Chapter 2).

The classification of genes into expression modules has revealed unexpected dynamics of genes, involved in a range of processes, previously not identified in the tapetum. An interesting example are the genes involved in gene silencing and chromatin organisation expressed early in pseudotime (section 3.4). Given the large-scale chromatin changes that occur during sexual reproduction [9, 152, 154, 161], and the tapetum's role as a source of small RNAs in grasses [156], this is an exciting hint that the tapetum may also undergo widespread changes in chromatin and DNA methylation through its development.

Parallel transcriptome and DNA methylome sequencing [263] would allow cells to be ordered in pseudotime by their transcriptomes, and changes in DNA methylation overlaid at the single-cell level. Given the early expression of Pol IV/V components, I hypothesise that early stages of tapetal development are times of large-scale changes in DNA methylation pattern.

Like many plant tissues, the tapetum undergoes endoreplication which gives rise to polyploid and multinucleate cells, most commonly with two tetraploid nuclei [16]. From pseudotime analysis several endoreplication-related genes were identified to be differentially expressed through tapetal development. *TOPOISOMERASE 6B* (*TOP6B*) is essential for endoreplication [278] and is expressed in the tapetum at a peak in mid-pseudotime. *INTERACTOR OF CDC2 KINASE* 4 (*ICK4*) blocks cell cycles progression [279] and increases in expression after *TOP6B* peaks. It is possible that these two modes of expression control the level of endoreplication in the tapetum. *TOP6B* (and other factors) may promote endoreplication at a specific stage, then *KRP6*/*ICK4* inhibits cell cycle progression. This would suggest that tapetum lacking a functional copy if KRP6/ICK4 would be of a higher ploidy level, or overproliferated. This data provides testable hypotheses for those exploring endoreplication.

As single-cells were collected in a way which preserves the genomic DNA while the mRNA is sequenced [262], it is possible to sequence the genomes of the same cells after they have been ordered along a pseudotime axis. This would allow me to investigate the progression of endoreplication through developmental time. Sequencing single-cells at different developmental time points would show if endoreplication occurs genome-wide in the tapetum, or at specific loci. The progression of endoreplication could be linked to gene expression at the single cell level, providing greater insight into the causes and consequences of endoreplication.

Discovery of novel tapetum expression profiles has led me to propose a small peptide signalling pathway functioning late in tapetal development. The late-stage (Module two) expression of *CLE1* and *RALFL1* suggests signalling from the tapetum to the microspores or pollen grains (Fig. 3-13). Analysis of these genes may reveal a novel signalling pathway controlling male germline development. A recent study has shown that small peptides may function in anthers to control pollen exine formation and patterning [280]. Expression of a dominant-negative CLE19 peptide results in deformed pollen exine, though it is worth noting that the *CLE19* gene studied is not expressed in the tapetum (Chapter 2). This suggests that the dominant-negative disrupts signalling of another small peptide.

3.8.3 Single cell sequencing can be used to infer gene regulatory interactions

While expression modules provide a broad view of gene expression, this can be extended to co-expression at the single cell level. Some genes, such as *LBD2* and *NAC10/SND3* show highly variable expression, even between cells close in pseudotime. At the single-cell level the expression of *LBD2* and *NAC10/SND3* was found to be strongly correlated, supporting a proposed regulatory interaction [218]. This highlights the type of analysis these data can be extended to. The discovery of genes co-expressed in single cells may provide greater knowledge about regulatory interactions and signalling. This may be particularly useful for genes that do not correlate with pseudotime, such as genes that regulate the cell cycle or

endoreplication. Pseudotime data could be further expanded to test gene regulatory networks, such as that proposed in Chapter 2, through weighted gene co-expression analysis [281]. Knowledge of co-expression in single-cells can help infer more reliable genetic interactions, as it is not confounded by variation between cells in the same tissue collected for bulk-sequencing.

3.9 Summary

While experiments on plant single cells are rare, this chapter has shown that they are not only feasible, but that the study of single cells can reveal novel properties of biological systems which are obscured through a bulk analysis. Overall, single-cell mRNA sequencing of the tapetum has proven to be a powerful approach to investigate gene expression and function in specific tapetal stages. Single-cell sequencing has revealed novel transcriptional dynamics which have not been previously identified. Pseudotemporal ordered tapetal transcriptomes should prove to be a valuable resource to plant researchers and provide greater insight into function of candidate genes affecting male fertility.

3.10 Methods

3.10.1 Single-cell isolation, library preparation, and sequencing

Tapetal cells from a *pA9::NTF* expressing line were protoplasted (sections 2.11.1 & 2.11.3) before being sorted using a BD FACS-Melody instrument (BD, Reading, UK). GFP-positive cells were sorted into individual wells of a 96-well plate. Plates were spun briefly at below 8000 *g* and frozen on dry ice. DNA and RNA were extracted, and cDNA was created with a modified Smart-seq2 [188] protocol according to [262]. Library preparation was performed using Nextera kits (Illumina). All libraries were pooled and sequenced on an Illumina Hiseq 2500 to a depth of 300M reads, with 100bp paired end reads.

3.10.2 Single-cell RNA-sequencing mapping

Galaxy [282] was used to create a reads-to-expression-matrix workflow. Reads from each single-cell library were mapped separately to the TAIR10 genome with RNA STAR [283]. FeatureCounts [284] was then used to measure gene expression in each BAM alignment file, the results of which were joined into a 96-well plate expression matrix, providing the expression for each TAIR10 gene in each single-cell. This expression matrix was then used for downstream QC and analysis.

3.10.3 Bulk RNA-sequencing analysis

Three WT bulk RNA-seq libraries (Appendix Table A1) and three plates of combined single-cell libraries were mapped to the TAIR10 genome using Tophat and Cufflinks [234]. Gene expression in fragments per kilobase per million reads was compared for all detected transcripts and correlation calculated using Pearson's product-moment correlation.

3.10.4 Monocle

Monocle [267] was implemented using gene expression in transcripts per million (TPM) for cells with more than 1500 genes detected and 25000 reads. Only single-cell libraries were used for analysis. Genes with a minimum expression of 0.1 TPM and expressed in 10 or more cells were used to infer the pseudotime trajectory. All other criteria used were default settings. Full scripts can be found at (https://github.com/Baldridge37/Monocle).

3.10.5 Manual pseudotime ordering

The same criteria were applied to filter cells as with the Monocle pipeline. Gene expression was normalised against the total number of genes expressed per cell using the method from Kieran Campbell (https://tinyurl.com/y8g3pogn). Expression of late-stage gene markers was found to correlate with the second principal component. Loadings of cells onto principal component two was used to plot raw expression

levels. The SC3 R package was used to cluster cells and extract marker genes for each cluster [273]. Full scripts can be found at (https://github.com/Baldridge37/Manual_singlecell).

3.10.6 SCORPIUS

SCORPIUS [270] was performed on all single cells without prior filtering of the data. Outlier cells were filtered based on their Pearson's correlation distances. A total of 183 single-cell libraries were used in the final analysis. The cell trajectory was inferred from the reduced dimensional representation of all cells, with no given number of clusters to discover. Gene expression modules were inferred for the 1000 most important genes for predicting the pseudotime ordering. Modules two, three, and four were combined as they differed only in their expression level and not their profile. Mean expression profiles were plotted as a loess fit of all genes in each expression module. Full code is available at (https://github.com/Baldridge37/SCORPIUS).

3.10.7 GO enrichment and plotting

GO term enrichment was calculated with Panther [191] and treemaps were created with REVIGO [285].

3.10.7 Pseudotime plotting

All pseudotime expression plots were created using ggplot2 [242]. Points represent expression in individual cells, and smoothed curves were created using local regression (LOESS), with 95% confidence intervals bounding the curve.

Chapter 4 The Tapetum methylome: an epigenetic bridge between soma and germline

4.1 Introduction

Methylation at the 5-carbon position of cytosine is an important modification made to the DNA of many eukaryotes. Cytosine DNA methylation affects a wide range of cellular processes, such as gene expression [286], splicing [9], meiotic crossovers [287], and most importantly, the silencing of transposons [288]. DNA methylation is measured in a number of ways, including mass-spectrometry, methylation-sensitive digestion, and localisation of methylated CG binding proteins. One of the most common, and the most comprehensive methods to assay cytosine methylation is whole genome bisulphite sequencing (BS-seq) [289]. Bisulphite treatment leads to the hydrolytic deamination of cytosine to uracil [289]. 5-Methylcytosine is protected from bisulphite treatment and so remains unmodified. When bisulphite-treated DNA is sequenced, the methylation level at a single cytosine, or in a genomic region, can be represented as the proportion of cytosines sequenced relative to the total of cytosines and converted thymines (from uracils) [289]. Throughout this chapter methylation level detected from BS-seq is measured as $\frac{Cytosines}{Cytosines + Thymines}$, where cytosines and thymines are the numbers of C or T bases present in a read, which maps to a genomic cytosine.

The major targets of cytosine DNA methylation in plants are transposons. DNA methylation across the length of transposons can be investigated through a method called ends analysis. Transposons are aligned at their 5′ and 3′ ends, and the methylation level across each half of the transposons averaged in windows of 100 base pairs (bp). In these plots the 5′ and 3′ ends represent short transposons and the edges of long transposons, while the centres show methylation at long transposons

(Fig. 4-1). Short, euchromatic transposons and transposon edges are usually targets of the RNA-directed DNA methylation pathway, while long, heterochromatic transposons are targeted by CHROMOMETHYLASE 2 (CMT2) and CMT3, through association with heterochromatic histone modifications [118] (Fig. 4-1). Through this method, one can infer the relative activities of these pathways between different cell types and mutants (Fig. 4-1).

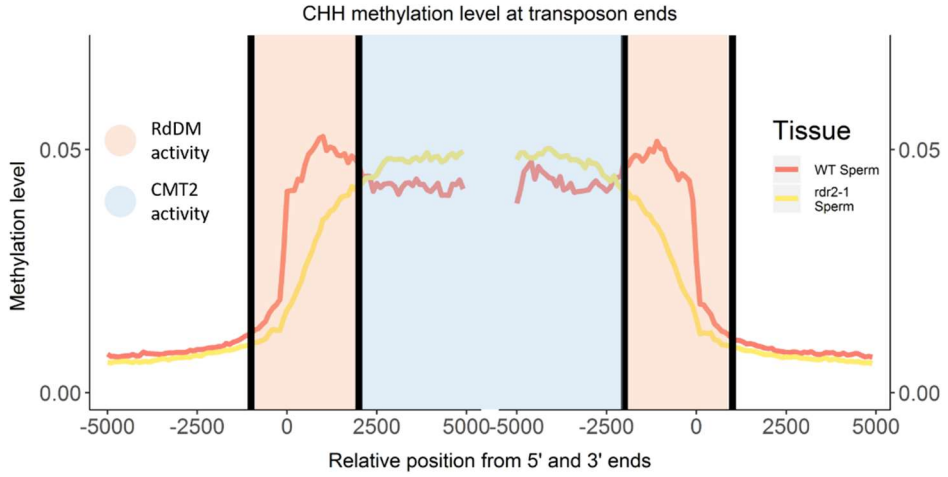


Figure 4-1: An example ends analysis plot showing the regions of transposons controlled by the RdDM pathway and the CHROMOMETHYLASES. RdDM acts at short euchromatic transposons and the edges of long heterochromatic transposons. Pol IV and V do not usually transcribe dense heterochromatin, and so methylation is maintained at long transposons through recruitment of the CMT methyltransferases to repressive histone marks. Ends analysis is a useful method to investigate the contributions of these methylation pathways in different cell types/mutants. In *rdr2* sperm, a reduction in methylation relative to WT is seen at transposons edges but not in heterochromatic centres.

The *Arabidopsis* male sexual lineage undergoes large scale chromatin rearrangements [163] and possesses hundreds of specific DNA hypermethylated sites which depend on the RdDM pathway [9]. In many organisms, germline-specific sRNAs and chromatin modifications are employed to control transposon activity during this important developmental window. In animals PIWI-interacting RNAs (piRNAs) direct chromatin modifications in the germline through sequence homology and in many animal systems piRNAs can be derived from somatic nurse cells to support germlines [173]. This reflects the proposed role of many plant nurse cells as a source of sRNAs for sexual lineage cells [139, 152, 154, 156].

This chapter aims to elucidate the DNA methylation profile of the male nurse cell layer, the tapetum, in *Arabidopsis thaliana*, and investigate the relationship between

tapetal and sexual lineage DNA methylation. I highlight novel features of the tapetum DNA methylome and explore the pathways responsible for this methylation. I experimentally test the hypothesis that the tapetum is a source of 24 nt sRNAs for the sexual lineage and propose a model to explain the differences in DNA methylation between the tapetum and sexual lineage, considering evidence of sRNA movement between the cell types.

4.2 Bisulphite sequencing reveals the tapetal methylome

Whole genome BS-seq of FACS-sorted tapetal cells has allowed the investigation of DNA methylation at single nucleotide resolution. Col-0 WT BS-seq data was mapped to the TAIR10 genome and methylation score calculated for each cytosine, and for cytosines in 50 bp windows (Appendix table A1). Three technical replicates of WT tapetum show strong correlation in 50 bp windows across the nuclear genome (average $R^2 = 0.866 \pm 0.006$, Pearson's product-moment correlation).

As transposons are the major target of DNA methylation, the level of methylation across transposons can be informative of gross differences between cell types and tissues. Transposon ends analysis of technical replicates of tapetum BS-seq data shows good agreement between them (data not shown). Ends analysis of transposons shows that the tapetum has somatic-level methylation in the symmetric CG context, and does not show the enhanced maintenance of methylation by METHYLTRANSFERASE 1 (MET1) that is seen in the sexual lineage [166] (Fig. 4-2). CHG methylation is lower than most sexual lineage cell types but is higher than somatic tissues such as seedling and rosette leaf, appearing at a similar level to sperm (Fig. 4-2). CHH methylation shows the biggest divergence in level and distribution across transposons between cell types. The tapetum appears to have sexual-lineage like CHH methylation, most like microspores (Fig. 4-2). Meiocytes and sperm possess low levels of CHH methylation across the length of all transposons and are below the level seen in the tapetum and other somatic cell types. The tapetum shows an enhanced level of methylation at transposon edges relative to rosette leaf and

seedlings, suggesting increased activity of the RdDM pathway (Fig. 4-2). This relationship is reversed toward the centre of long transposons, with the tapetum showing lower methylation levels than seedlings or rosette leaves (Fig. 4-2). Together with the lack of CMT2 expression in the tapetum (RNA-sequencing), this suggests a reduction in heterochromatic DNA methylation.

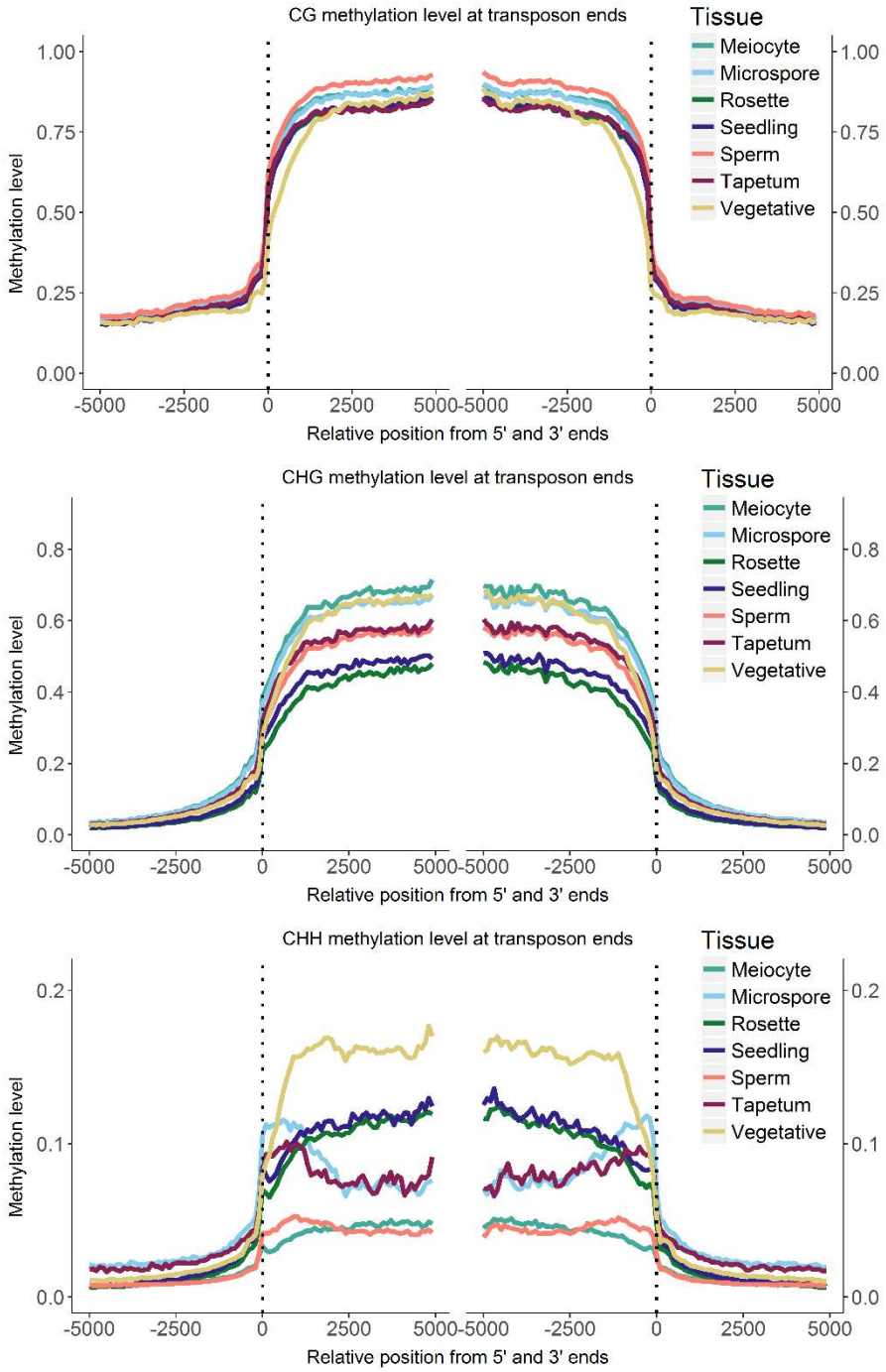


Figure 4-2: Transposon ends analyses of meiocyte, microspore, rosette leaf, seedling, sperm, tapetum and vegetative cell BS-seq data. The tapetum appears to possess intermediate transposon methylation patterns between the sexual lineage and somatic cells. Tapetum transposon methylation most closely matches that seen in microspores and shows higher CHH methylation at transposon edges than other somatic tissues (rosette leaf and seedling), suggesting an enhanced action of RdDM.

Unlike transposons, genes largely lack DNA methylation, except in the CG context in the bodies of around a third of genes [290]. From gene ends analysis we can see that the tapetum does not differ significantly from other cell types and tissues in terms of gene body methylation (Fig. 4-3).

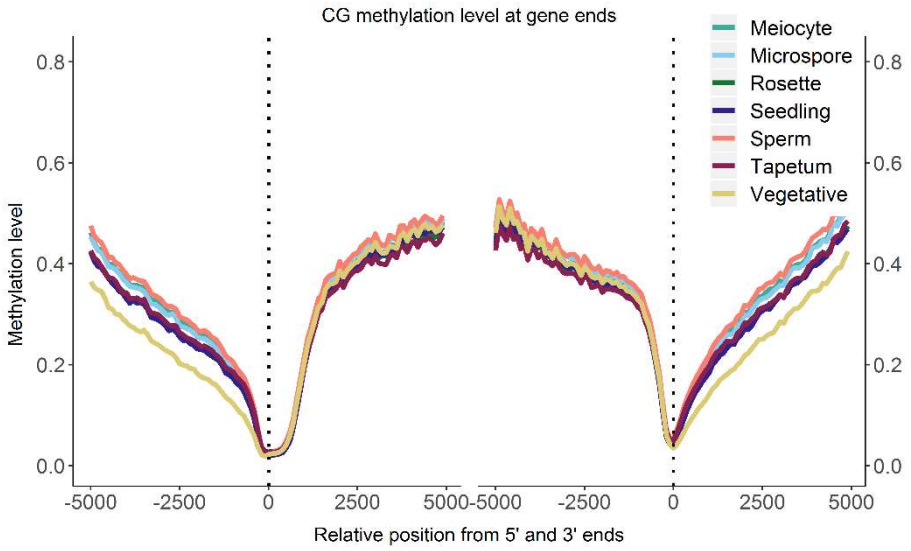


Figure 4-3: Gene ends analyses of meiocyte, microspore, rosette leaf, seedling, sperm, tapetum and vegetative cell BS-seq data. The tapetum does not appear to differ from other cell types with regards to CG methylation of genes.

Given the strong methylation of transposon edges (Fig. 4-2), the contribution of the RdDM pathway was investigated by collecting tapetal cells from an *rdr2-1* mutant (referred to as *rdr2*). Three replicate libraries of *rdr2* showed good correlation in 50 bp windows (mean R^2 = 0.737±0.027, Pearson's product-moment correlation) and agreement in ends analysis (data not shown, Appendix table A1). Transposon ends analysis shows that *rdr2* tapetum has lower methylation across transposons in all contexts (Fig. 4-4). While only a small difference in CHH methylation can be seen, it is interesting to note that for both CG and CHG methylation, lower methylation is apparent across the length of long transposons (Fig. 4-4). This contrasts with RdDM activity in other tissues, where only transposons edges show differences in methylation in RdDM-mutants (Figs. 4-1 & 4-4) [127]. To understand the methylation of *rdr2* tapetum genome-wide, I further examined RdDM-target loci in the sexual lineage and soma [9]. I found both groups of RdDM loci are significantly less

methylated in *rdr2* tapetum than WT tapetum in all sequence contexts ($p < 0.001$, Conover-Iman test) (Fig. 4-5), confirming results from ends analysis.

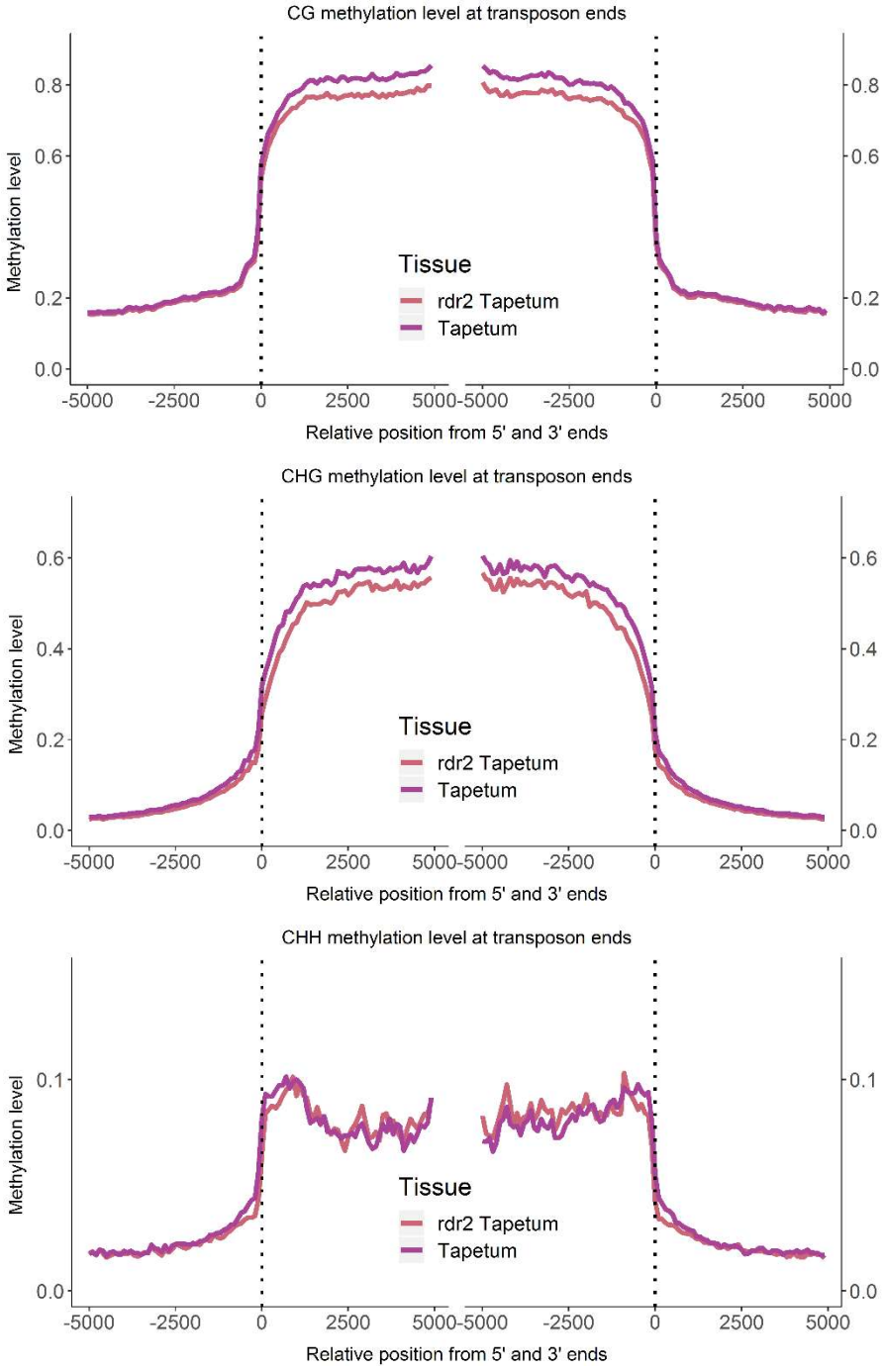


Figure 4-4 Transposon ends analysis of WT and *rdr2* tapetum bisulphite-sequencing data. Methylation is lower in *rdr2* in all sequence contexts, though this is only clear at transposon edges for the CHH context. Interestingly CG and CHG methylation are lower across the length of transposons, rather than just at the transposon edges, which are canonical targets of RdDM.

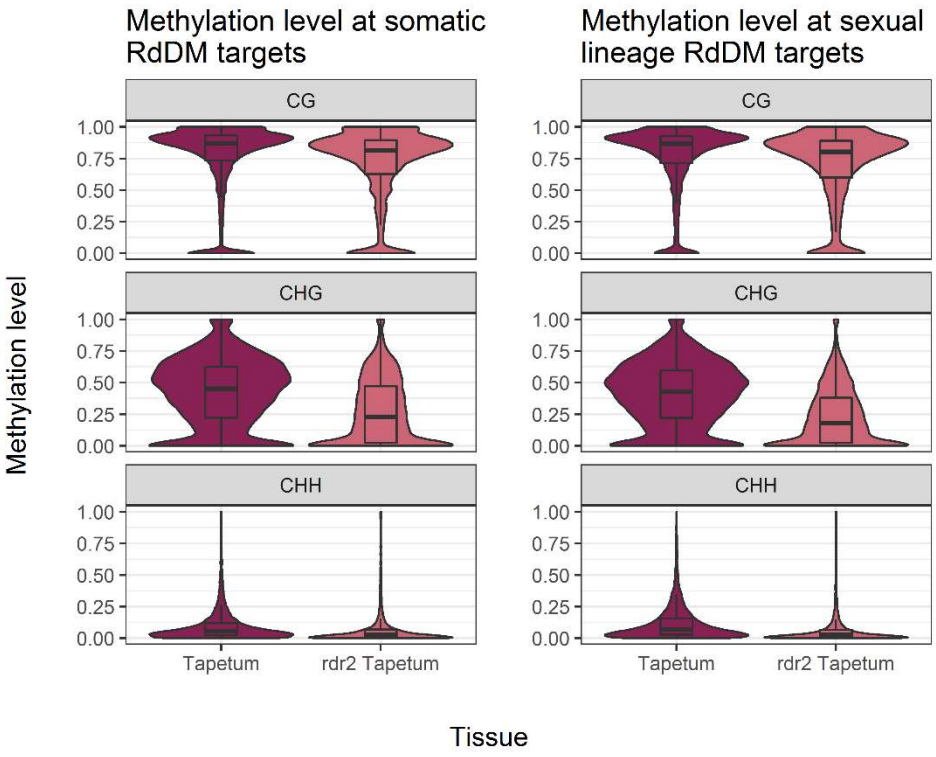


Figure 4-5 Cytosine methylation levels in WT and *rdr2* tapetum at somatic or sexual lineage RdDM targets. In all contexts *rdr2* tapetum shows lower levels of methylation than WT.

4.3 The tapetum shows regions of hypomethylation relative to somatic tissues

Active demethylation of DNA occurs in the companion cells of gametes, the male vegetative cell and the female central cell. This demethylation reinforces silencing of transposons in the germline via sRNAs [152, 154]. As *DEMETER* (*DME*) is expressed in the tapetum (RNA-seq), whether the tapetum also shows active DNA demethylation was investigated.

DNA methylation in the tapetum was compared to somatic cell types (seedlings, rosette leaves, cauline leaves and roots) to find hypomethylated regions. Differentially methylated regions (DMRs) were defined as regions with significantly lower CHH methylation than the somatic cell types, and also showed significantly lower CHG, or CG methylation (Methods 4.10.7). In total there are 971 sites significantly hypomethylated in the tapetum relative to somatic tissues ($p < 0.001$, Fisher's exact test). Hypomethylated DMRs are short, with an average length of 302

bp. Given the expression of *DME* in the tapetum, and DME's function in the vegetative cell, I investigated if hypomethylated sites could be DME targets. A total of 12116 DME-target sites in the vegetative cell have previously been identified (unpublished data). I found that 349 hypomethylated DMRs (35.9%) overlap with vegetative-cell DME-target sites.

To assess the significance of this association with DME-target sites, lists of random control loci were generated with the same length distributions as hypomethylated DMRs. Lists were also controlled for the proximity of loci to either genes or transposons (Table 4-1). There was a significantly lower number of loci overlapping vegetative DME-target sites for all control lists than for hypomethylated DMRs ($p < 0.001$, Fisher's exact test). This suggests that tapetum hypomethylated DMRs are more likely to overlap vegetative cell DME-target loci than similarly located sequences in the genome

List	Non-overlapping	Overlapping DME-targets	Proportion overlapping	*p* value†
Hypomethylated DMRs	622	349	35.94%	
Gene control 1	859	112	11.53%	$p = 4.60 \times 10^{-38}$
Gene control 2	850	121	12.46%	$p = 8.48 \times 10^{-35}$
Gene control 3	859	112	11.53%	$p = 4.60 \times 10^{-38}$
Transposon control 1	868	103	10.61%	$p = 1.39 \times 10^{-41}$
Transposon control 2	869	102	10.50%	$p = 5.43 \times 10^{-42}$
Transposon control 3	882	89	9.17%	$p = 1.26 \times 10^{-47}$
Somatic RdDM	4350	5466	55.68%	$p = 1.83 \times 10^{-32}$
Sexual lineage RdDM	3488	5563	61.46%	$p = 1.20 \times 10^{-52}$

Table 4-1: DMR lists and their overlaps with vegetative cell DME-targets. Hypomethylated DMRs show a significantly greater proportion of overlap than control lists. Somatic and sexual lineage RdDM-targets show a significantly greater overlap with vegetative cell DME-targets, than hypomethylated DMRs. †All *p*-values are from Fisher's exact test in comparison to hypomethylated DMRs.

As DME-targets are known to overlap with RdDM-targets [152], I examined the DNA methylation at hypomethylated DMRs in WT and *rdr2* tapetum to understand whether hypomethylated DMRs are also subject to RdDM control (Fig. 4-6). Methylation in all contexts is significantly lower in *rdr2* tapetum than WT ($p < 0.001$, Conover-Iman test), demonstrating that these sites are also targets of RdDM (Fig. 4-6). While hypomethylated DMRs were defined by comparison between the tapetum and somatic tissues, meiocytes also show significantly lower methylation in the CHG

and CHH contexts than rosette leaf ($p < 0.001$, Conover-Iman test, Fig. 4-6). I then examined the proportion of sexual lineage and somatic RdDM loci that overlap vegetative cell DME-targets respectively. Both sets of RdDM-targets overlap vegetative cell DME-targets at a significantly higher rate than tapetum hypomethylated DMRs (Table 4-1), showing that RdDM activity correlates with DME activity in the vegetative cell.

While it is possible that hypomethylated DMRs are targets of DME in the tapetum, there is no direct evidence to support this. As RdDM-targets show a significantly higher overlap with DME-targets, and RdDM controls methylation at hypomethylated DMRs, it is plausible that the overlap with DME-targets is through an indirect association. Hypomethylated DMRs show significantly lower methylation in meiocytes than rosette leaf ($p < 0.001$, Conover-Iman test, Fig. 4-6), suggesting that hypomethylated DMRs are RdDM-targets of low activity in anther tissues.

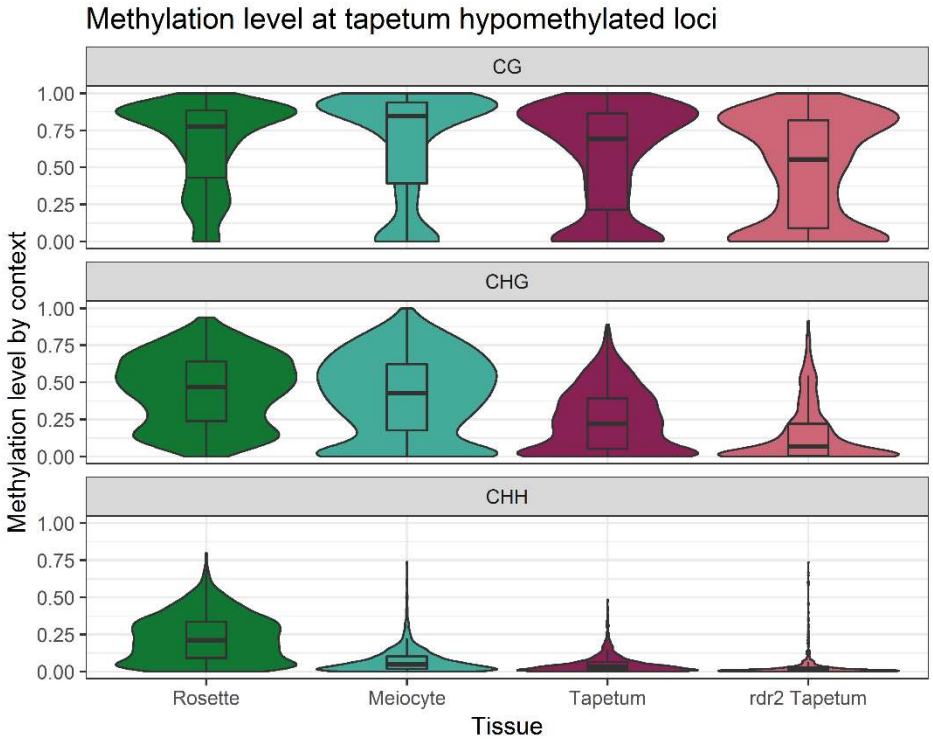


Figure 4-6 : Cytosine methylation levels at tapetum hypomethylated loci. Hypomethylated loci show high methylation in rosette leaf but low methylation in the tapetum and meiocytes. Loci show dependence on RDR2, suggesting they are still targets of RdDM.

4.4 The tapetum shows hypermethylation at specific loci

Comparing tapetal DNA methylation to that of somatic tissues (seedlings, rosette leaves, cauline leaves and roots) revealed sites that are strongly hypermethylated in the tapetum. DMRs were selected if they showed significantly higher CHH methylation in the tapetum compared to somatic tissues, as well as significantly higher CHG or CG methylation (Methods 4.10.7). This yielded a total of 2549 hypermethylated DMRs with significantly different methylation in the CHH and, CG or CHG, context. Of these, 1713 loci (67%), are significantly hypermethylated in WT tapetum compared to *rdr2* mutant, and hence are dependent on RDR2 ($p < 0.001$, Fisher's exact test). These RDR2-dependent tapetum-hypermethylated DMRs (simplified as tapetum RDR2 DMRs) are distributed throughout the genome and have an average length of 372 bp.

A substantial proportion (413; 24%) of tapetum RDR2 DMRs overlap canonical sexual-lineage-hypermethylated loci (SLHs, 365 out of a total of 724 SLHs, Figs. 4-7 & 4-8). A further 31 tapetum RDR2 DMRs overlapped sites of sexual-lineage-specific methylation (SLMs, 24 out of a total of 469 SLMs), which are methylated *de novo* in the sexual lineage (Figs. 4-7 & 4-8). While these tapetum RDR2 DMRs overlap sites hypermethylated in the sexual lineage, it is worth noting that the vast majority (95.5%; 424 out of 444) possess stronger methylation in the tapetum than the sexual lineage (Fig. 4-7).

To find tapetum-specific RDR2 DMRs, sites overlapping either SLMs or SLHs were removed. This produced 1269 loci, which I called tapetum-specific hypermethylated DMRs, for further study. Tapetum-specific hypermethylated DMRs have an average length of 322 bp. As methylation at SLHs and SLMs was found to be entirely dependent on RDR2 [9], I sought to investigate whether tapetum-specific hypermethylated DMRs are also solely RDR2-dependent. Bisulphite-sequencing of *rdr6* (*sgs2-1*) mutant tapetum was performed and three biological replicates of *rdr6* mutant tapetum showed good agreement in transposon ends analysis, and

correlation in 50 bp windows (mean R^2 = 0.660±0.005, Pearson's product-moment correlation, Appendix table A1). Methylation at tapetum-specific hypermethylated DMRs was found to be significantly reduced in the *rdr6* mutant compared to WT (p < 0.001, Conover-Iman test, Fig. 4-9). The reduction in methylation shows that RDR6 is also involved in the hypermethylation of these loci. As expected, *rdr2* tapetum showed significantly lower methylation in all sequence contexts compared to WT (p < 0.001, Conover-Iman test, Fig. 4-9), demonstrating that both RDR2 and RDR6 control DNA methylation at these sites.

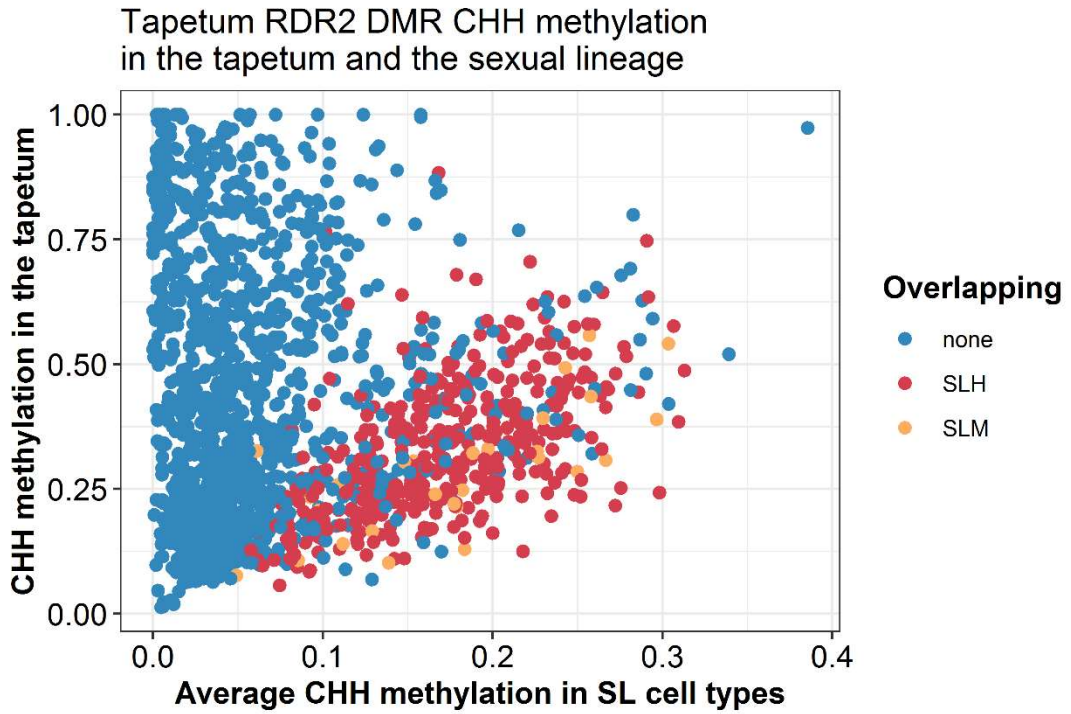


Figure 4-7: Methylation in the CHH context at tapetum RDR2-DMRs in the sexual lineage (SL) and tapetum. CHH methylation at most tapetum RDR2-target loci is low in the sexual lineage, but notable exceptions are those that overlap SLMs and SLHs. Sexual lineage average methylation represents the methylation score of the combined data of meiocytes, microspore, and sperm.

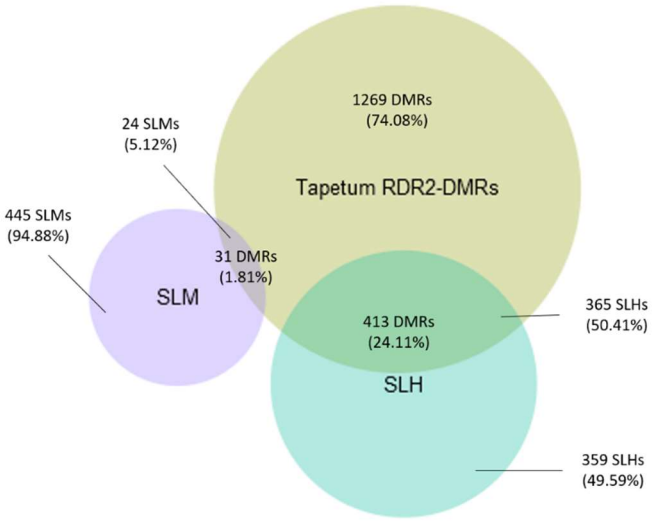


Figure 4-8: Overlap of Tapetum RDR2-DMRs, SLMs, and SLHs. A total of 444 Tapetum RDR2-DMRs overlap either SLMs or SLHs, leaving 1269 tapetum-specific DMRs. 413 Tapetum RDR2 DMRs overlap 365 SLHs, and 31 Tapetum RDR2 DMRs overlap 24 SLMs. The overlap with SLHs (50.41% of SLHs) is far greater than the overlap with SLMs (5.12% of SLMs).

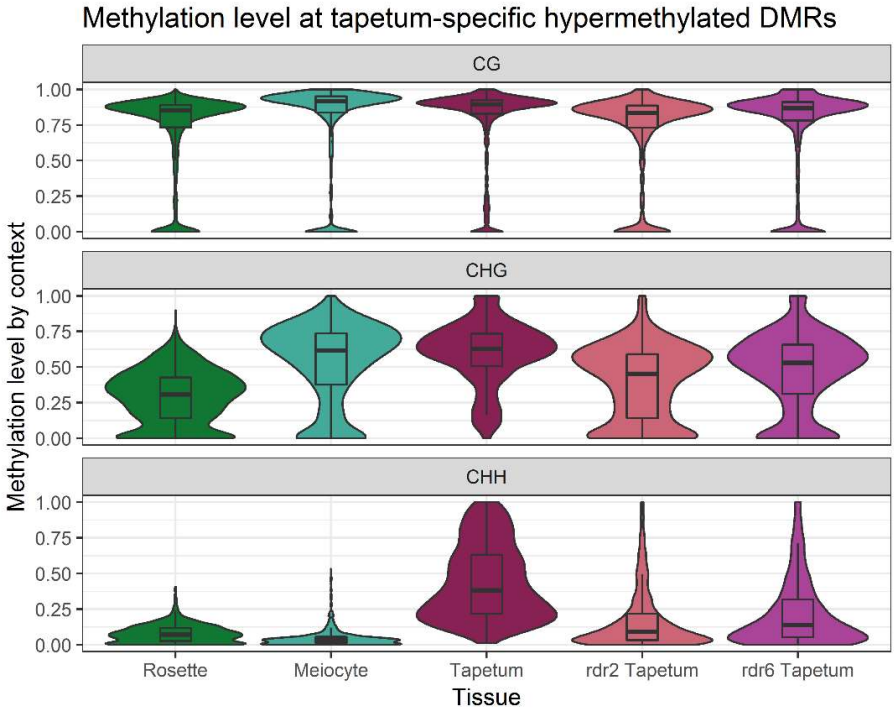


Figure 4-9: Cytosine methylation levels at tapetum-specific hypermethylated DMRs. DMRs show methylation in other somatic cell types (rosette leaf) suggesting the hypermethylation in the tapetum does not occur *de novo*. Methylation at these loci shows dependence on both RDR2 and RDR6.

While average methylation at tapetum-specific hypermethylated DMRs is significantly lower in both the *rdr2* and *rdr6* mutants, it is unclear if all DMRs are affected equally, or if distinct groups constitute these DMRs. Comparing methylation at individual sites shows DMRs do not form distinct clusters of RDR2- or RDR6-depedence, but rather form a continuum, with the majority of DMRs depending on both polymerases (Fig. 4-10). Indeed, methylation is highly correlated, but with a greater role of RDR2 than RDR6 (Fig. 4-10) (R^2 = 0.71, Pearson's product-moment correlation).

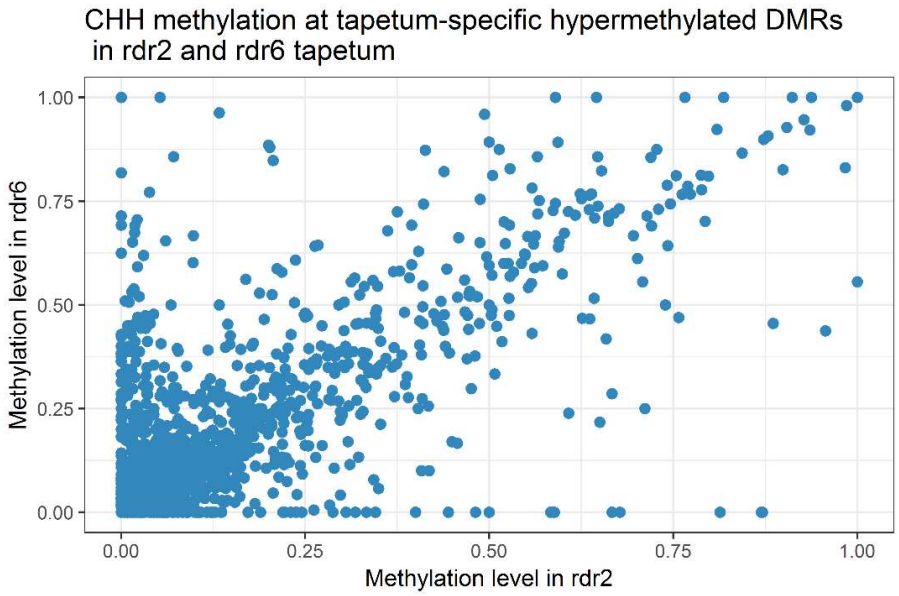


Figure 4-10: CHH methylation level at tapetum-specific hypermethylated DMRs in rdr2 and rdr6 tapetum. Methylation in both mutants is highly correlated, suggesting that these loci are dependent on both RDR2 and RDR6 for methylation. R^2 = 0.71, Pearson's product-moment correlation.

Tapetum-specific hypermethylated DMRs show a similar association (within 500 bp) with transposons as do RdDM-targets, though are less likely to overlap transposons (Fig. 4-11). Unlike SLMs, tapetum-specific hypermethylated DMRs do not show an increased association with genes, rather than transposons (Fig. 4-11)[9]. A total of 418 genes are within 500 bp of tapetum-specific hypermethylated DMRs, and 249 genes overlap these sites. There are no significantly enriched Gene Ontology terms associated with either overlapping genes, or those within 500 bp [189-191]. As methylation controlled by RdDM can play important roles in gene regulation [9, 286],

the effect of DMR methylation on gene expression was investigated through RNA-sequencing of *rdr2* tapetum.

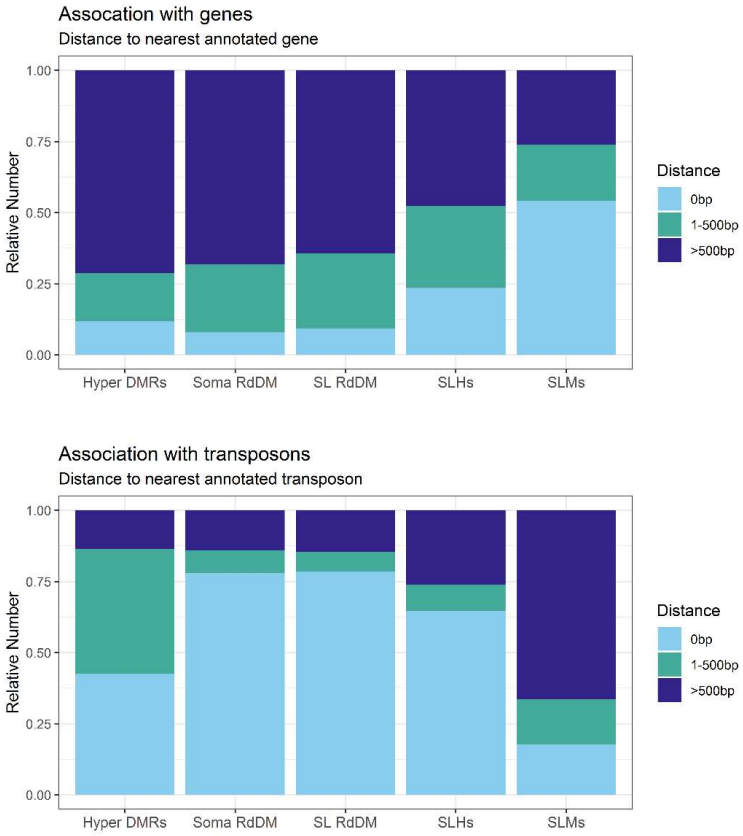


Figure 4-11: The genomic localisation of tapetum-specific hypermethylated DMRs (Hyper DMRs), somatic and sexual lineage RdDM-targets (Soma RdDM and SL RdDM), SLHs and SLMs [9]. Tapetum-specific hypermethylated DMRs show a similar association with genes as RdDM-targets, and do not show an increased association with genes like SLMs. Tapetum-specific hypermethylated DMRs show a similar proportion of DMRs within 500 bp of transposons as RdDM targets, though with fewer overlapping. Distances represent the distance to the nearest gene or transposons from each DMR/RdDM target.

Of the 418 genes within 500 bp of tapetum-specific hypermethylated DMRs, 31 (7.4%) are significantly upregulated in the *rdr2* mutant ($p < 0.05$, Cuffdiff.). This proportion does not differ significantly from the 8.9% of all loci tested (1656 out of 18633) that were found to be significantly upregulated ($p = 0.372$, χ^2 test). For downregulated genes the situation is similar, with the proportion of downregulated genes associated with DMRs (6; 1.5%) not differing significantly from the proportion genome wide (567, 3.0%) ($p = 0.067$, χ^2 test). These results suggest that these tapetum-specific hypermethylated DMRs are not more associated with differentially expressed genes than chance and hence do not preferentially regulate gene expression in the tapetum,

though specific genes may show regulation by tapetum-specific hypermethylated DMRs (Appendix, table A2).

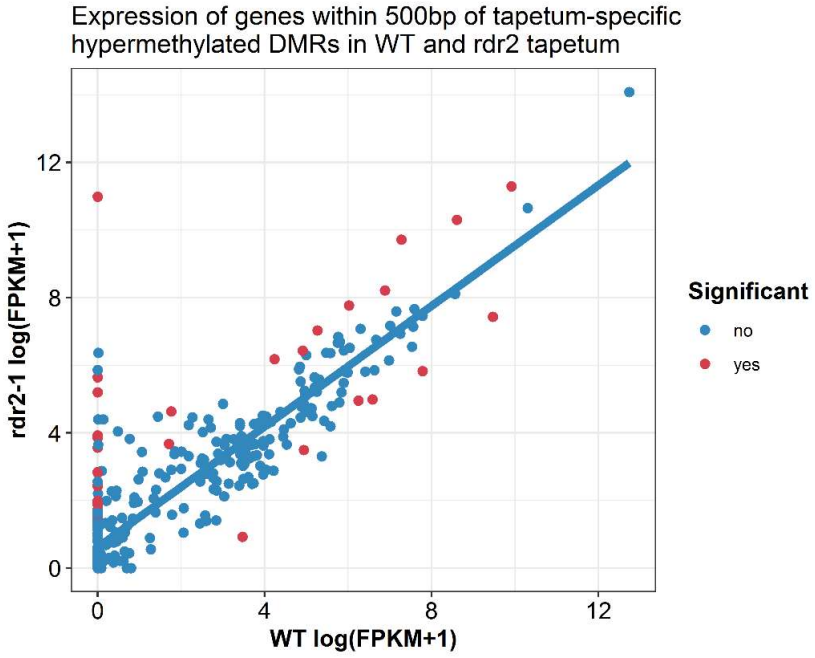


Figure 4-12 Expression of genes within 500 bp of tapetum-specific hypermethylated DMRs in WT and rdr2 tapetum. Red points show genes that are significantly differentially expressed between the two genotypes (Cuffdiff, $p < 0.05$). Genes associated with tapetum-specific hypermethylated DMRs are not more likely to be differentially expressed between WT and rdr2 than chance. Line of best fit follows the formula Log2[rdr2] = 0.8922*Log2[WT] + 0.6166

4.5 The Tapetum possesses strong DNA methylation at SLHs, but not SLMs.

In *Arabidopsis thaliana*, the male sexual lineage shows DNA hypermethylation at specific loci. SLHs were found to retain low levels of CG/CHG methylation in somatic tissues, while SLMs were re-methylated *de novo* in the male sexual lineage at each generation [9]. Like canonical RdDM targets, SLHs were found to be mainly associated with transposons, though SLMs showed an increased association with genes (Fig. 4-11), and play important roles in regulating their expression [9].

As described in Section 4.4, tapetum RDR2 DMRs overlap a substantial proportion of SLHs, and some SLMs (Figs. 4-7 & 4-8). To comprehensively understand the methylation patterns over all SLHs and SLMs, methylation in wild type, *rdr2*, and

rdr6 tapetum was investigated. While SLHs retain methylation in somatic tissues [9], their methylation pattern in the tapetum closely matches that seen in the meiocyte, and the two cell types are not significantly different in the CHH context (p = 0.086, Conover-Iman test) (Figs. 4-13 & 4-14). Consistently 480 out of 724 (66.3%) SLHs have significantly higher CHH methylation and CHG (or CG) methylation in the tapetum than rosette leaf (p < 0.001 in each context, Fisher's exact test), demonstrating that SLHs are predominantly hypermethylated in the tapetum. As in the sexual lineage [9], methylation at SLHs in the tapetum was shown to be largely dependent on RDR2 but not RDR6 (Figs. 4-13 & 4-14). This contrasts with tapetum-specific hypermethylated loci that show a reduction in methylation in both *rdr2* and *rdr6* mutants (Figs. 4-9 & 4-10).

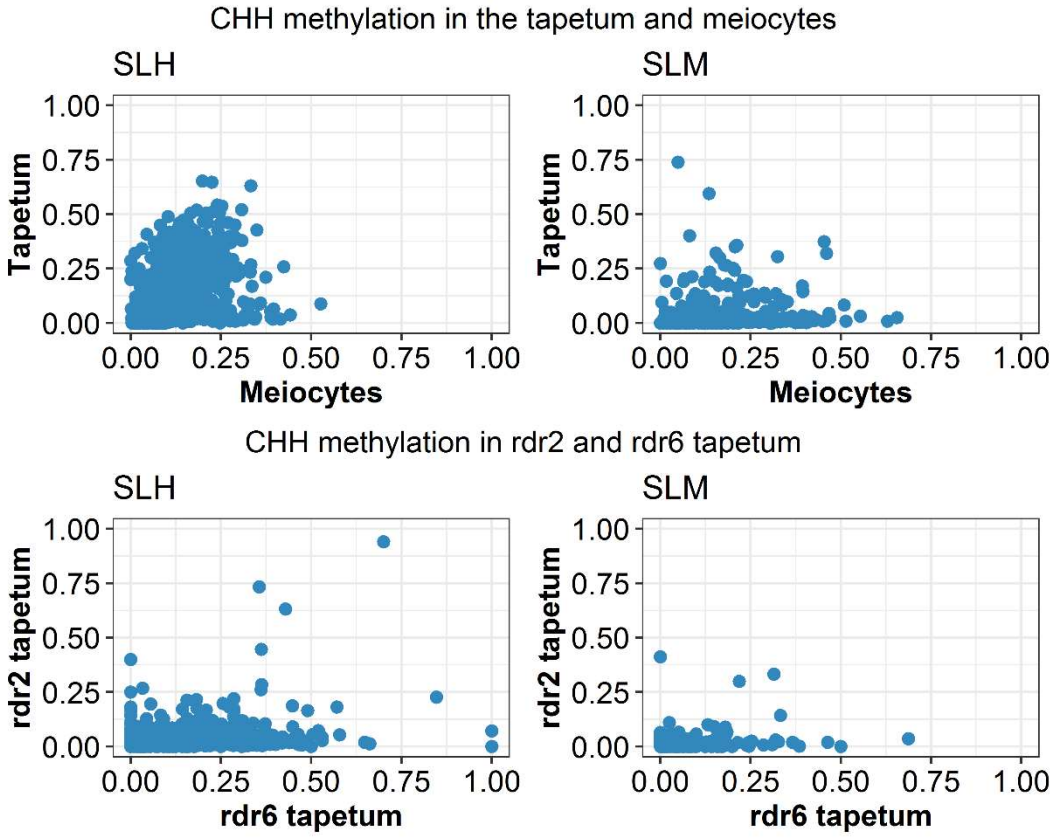


Figure 4-13: Scatter plot showing CHH methylation levels at individual SLHs or SLMs in the tapetum and meiocytes, and *rdr2* and *rdr6* tapetum. SLHs show strong methylation in the tapetum and methylation is somewhat correlated between the tapetum and meiocytes R^2 = 0.36, Pearson's product-moment correlation. The majority of SLMs appear to only be methylated in meiocytes and methylation between the two cell types is not correlated. Like in the sexual lineage, SLH and SLM methylation depends only on RDR2. Methylation in *rdr6* tapetum appears WT at SLHs in the tapetum.

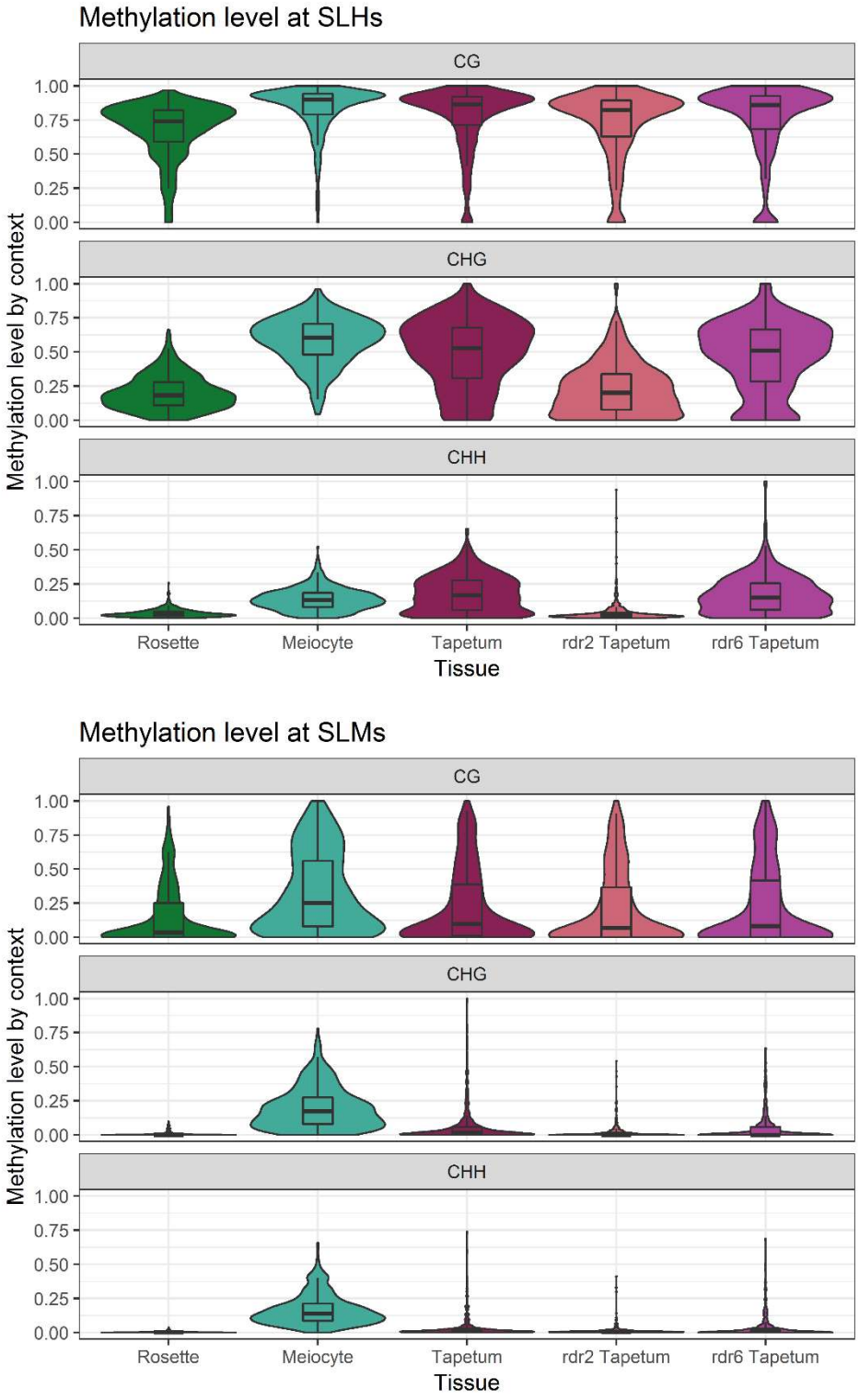


Figure 4-14 Methylation level at SLHs and SLMs in rosette leaf, meiocytes, and WT, *rdr2*, and *rdr6* tapetum. The tapetum shows strong methylation at SLHs but has very low levels of methylation at SLMs. Methylation at SLHs depends only on the canonical RdDM-pathway and not on RDR6. The tapetum appears intermediate between the soma and sexual lineage, with high SLH methylation like the sexual lineage, but low methylation at SLMs like somatic tissues.

At SLMs, however, methylation in all contexts in WT tapetum was far below the level seen in meiocytes ($p < 0.001$, Conover-Iman test, Figs. 4-13 & 4-14). Only 133 out of 469 (28.3%) SLMs have significantly more CHH methylation and CHG (or CG) methylation in tapetum than rosette leaf ($p < 0.001$ in each context, Fisher's exact test). This demonstrates that hypermethylation at SLMs is largely a specific feature of the sexual lineage and is not present in the tapetum.

4.6 SLM methylation is induced by SLH-derived sRNAs binding with mismatches

To investigate RdDM activity in the sexual lineage, small RNA libraries were created from WT meiocyte cells (unpublished data). When mapped to the *Arabidopsis thaliana* genome, many 24 nt sRNAs were mapped to SLHs (Fig. 4-15). However, when zero mismatches were allowed in mapping (i.e. perfect base pairing of sRNAs), very few 24 nt sRNAs were found to map to SLMs (Fig. 4-15). In fact, the majority (262/469; 55.9%) of SLMs show no sRNA mapping, but still show *de novo* methylation in the meiocytes. When up to three mismatches were allowed when mapping sRNAs, there was no significant difference between the sRNA reads (weighted RPKMs) at SLHs relative to zero mismatching ($p = 0.096$, Mann-Whitney U test). At SLMs however there is a large and significant increase in small RNAs mapping ($p < 0.001$, Mann-Whitney U test) (Fig. 4-15).

sRNAs mapping to SLMs with mismatches were further investigated and found to also map perfectly to SLHs. This suggests that SLHs may be a source of 24 nt sRNAs to direct methylation at SLMs. When up to three mismatches are allowed, 8580 best-hit sRNA reads were mapped to SLMs, in an SLH-masked genome. Of these sRNA reads, 1037 (12%) were found to map to SLHs perfectly (unpublished data).

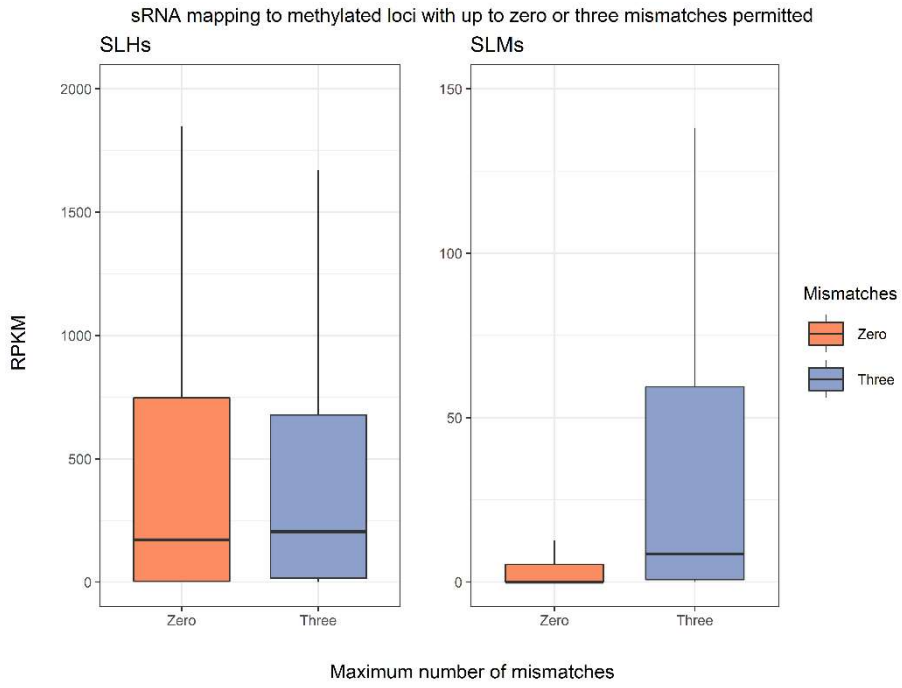


Figure 4-15: Small RNAs mapping to SLM and SLHs in meiocytes. While SLHs show many small RNAs mapping without mismatches, SLMs show very low levels of sRNA mapping. Allowing sRNAs to map with mismatches and taking the weighted RPKM has little effect on the total pool of sRNAs mapping to SLHs but massively increases those mapping to SLMs. Many sRNAs mapping to SLMs with mismatches map to SLHs perfectly. Boxplots do not show outliers and lines represent 1.5 IQR.

4.7 The tapetum can be a source of small RNAs to direct DNA methylation in meiocytes

Meiocytes show hypermethylation at SLMs but lack perfectly matching 24 nt sRNAs, showing that SLMs are not 24 nt sRNA source loci (Fig. 4-15). This suggests a decoupling of Pol IV and Pol V pathways in meiocytes, implying that the surrounding tapetal cells are a source of sRNAs for the meiocytes. This would fit with the role of the tapetum as a source of 24 nt phasiRNAs in grasses [156]. To test this hypothesis, RDR2-complementation transgenic lines expressing a FLAG-tagged RDR2 specifically in the tapetum in an *rdr2* background were created (Fig. 4-16). The promoter *pA9* has previously been used to drive expression in the tapetum [34], and shows strong and tapetum-specific expression of RDR2-FLAG protein (Fig. 4-16). Meiocytes were extracted to investigate DNA methylation by bisulphite sequencing.

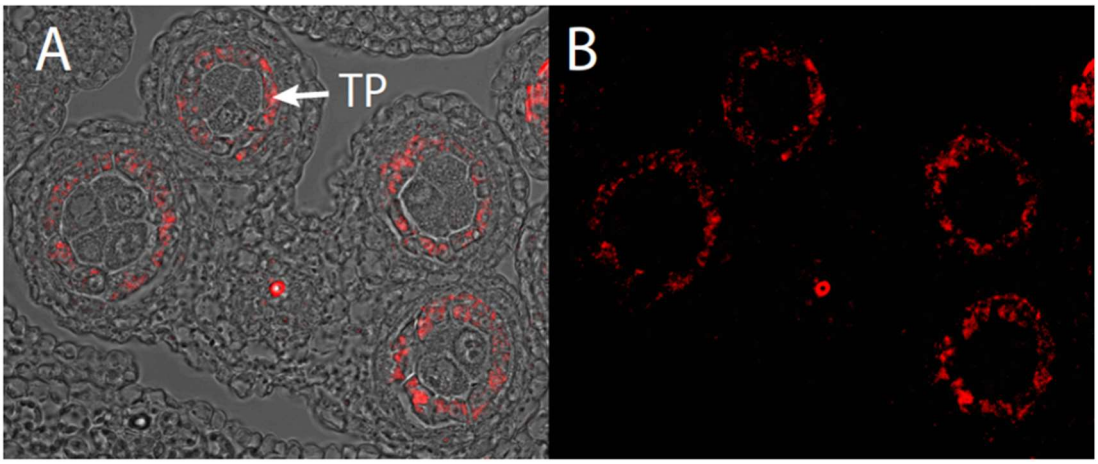


Figure 4-16: Immunostaining using α-FLAG on *pA9::RDR2:FLAG rdr2* #2-4 anthers. Strong and specific RDR2 expression can be seen in the tapetal cells and RDR2 expression is absent from meiocytes. A- Brightfield and fluorescence overlay, B- Fluorescence signal only. TP-Tapetum.

To test whether tapetal expression of RDR2, in an *rdr2* mutant background, could rescue methylation in meiocytes I investigated DNA methylation at known-methylated sites. For each sequence context (CG, CHG, and CHH) 50 bp windows that are strongly methylated in rosette leaf were defined. Two independent transformant lines of *pA9::RDR2 rdr2*, #1-3 and #2-4, were investigated for their recovery of methylation in these windows. Meiocytes from *pA9::RDR2 rdr2* lines #1-3 and #2-4 were compared to WT and *drm1;drm2* (Fig. 4-17), and sperm from *pA9::RDR2 rdr2* line #2-4 was compared to WT and *rdr2* (Fig. 4-17). In both meiocytes and sperm CG and CHG methylation levels recovered to similar levels, between 91% and 99% of WT (Fig. 4-17). A marked difference was seen in CHH methylation where, in all replicates, methylation levels were above 100% of WT levels (Fig. 4-17). This ranged from 108% in *pA9::RDR2 rdr2* #2-4 sperm, to 128% in *pA9::RDR2 rdr2* #1-3 meiocytes (Fig. 4-17). As *pA9::RDR2 rdr2* are expressing 24 nt sRNAs in the tapetum this recovery of methylation in meiocytes must be brought about by the DRM methyltransferases in the Pol-V pathway [128, 291]. The increased recovery of methylation at CHH than CHG or CG sites is consistent with a preference of the DRM methyltransferases for CHH sites [142, 292]. As RDR2 is expressed only in the tapetum in the *rdr2* mutant background, and re-methylation must occur each generation, these transgenic lines could prove to be a valuable system to follow the dynamics of RdDM methylation *in planta*.

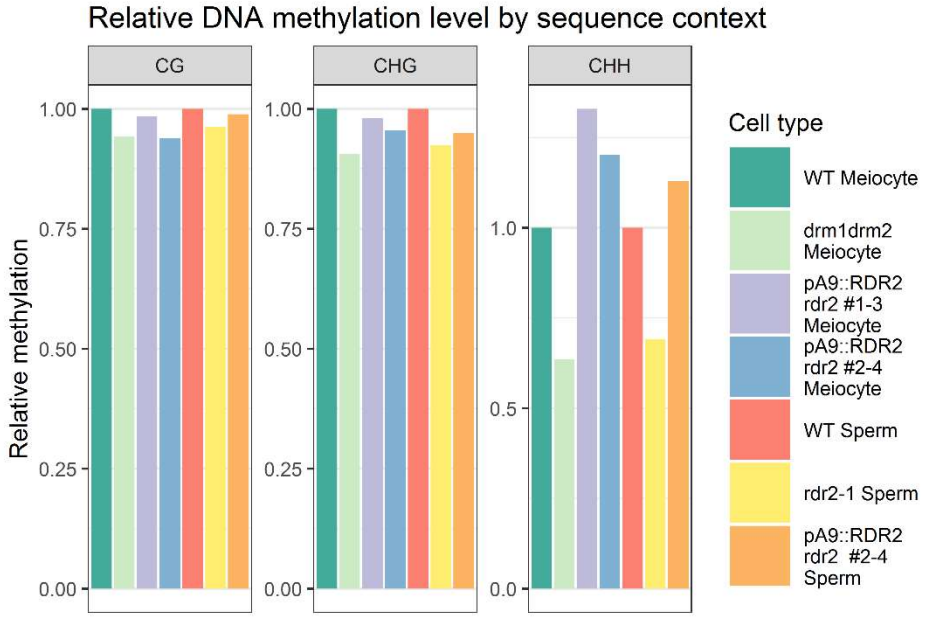


Figure 4-17: DNA methylation levels relative to WT in *pA9::RDR2 rdr2* meiocytes and sperm in highly methylated 50 bp windows. In all meiocyte and sperm *pA9::RDR2 rdr2* lines CG and CHG methylation remains below WT levels but CHH methylation recovers to above WT levels, showing the sequence preference of the DRM methyltransferases.

To explore methylation recovery at transposons, ends analysis was performed comparing WT, *pA9::RDR2 rdr2*, and RdDM-mutant meiocytes and sperm. In *pA9::RDR2 rdr2* lines, CHH methylation has recovered above WT levels, particularly at transposon edges which are targets of RdDM activity (Figs. 4-18 & 4-19). CHH methylation in WT and *pA9::RDR2 rdr2* #2-4 sperm reach similar levels at the centre of long transposons, which are heterochromatic and targeted by CMT2, and as such are less influenced by RdDM (Fig. 4-18). However, methylation in the CHH context is higher in *pA9::RDR2 rdr2* meiocytes across the full length of transposons compared to WT (Fig. 4-19).

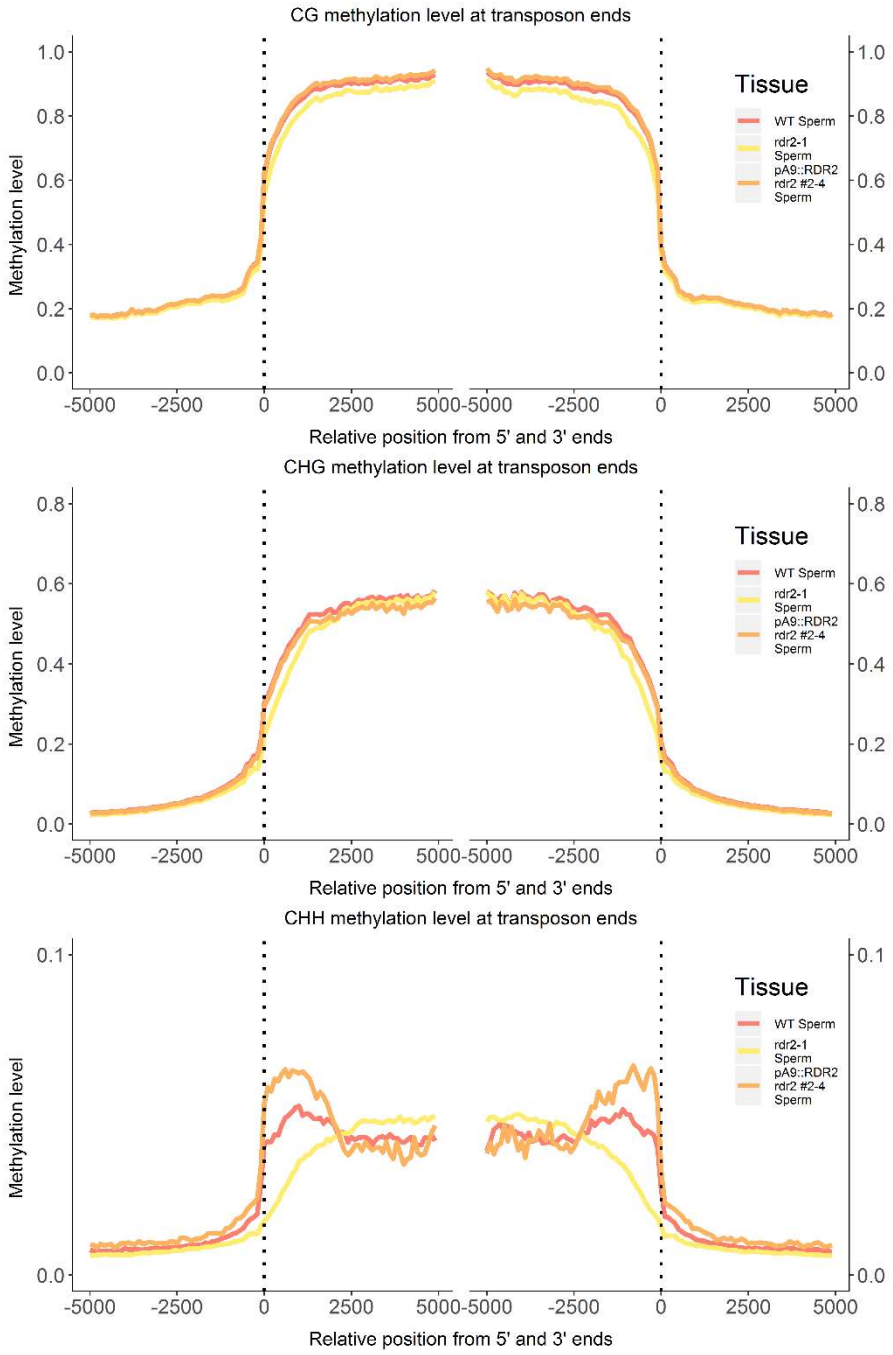


Figure 4-18: Transposon ends analyses of *pA9::RDR2 rdr2* sperm (#2-4) with WT, and *rdr2* sperm. Methylation in the CG and CHG contexts recovers to WT levels or higher in all *pA9::RDR2 rdr2* lines. In the CHH context, methylation at transposon edges is higher than WT levels, showing a clear role of RdDM at these sites. This reduces to WT levels in over long heterochromatic transposons in *pA9::RDR2 rdr2* sperm.

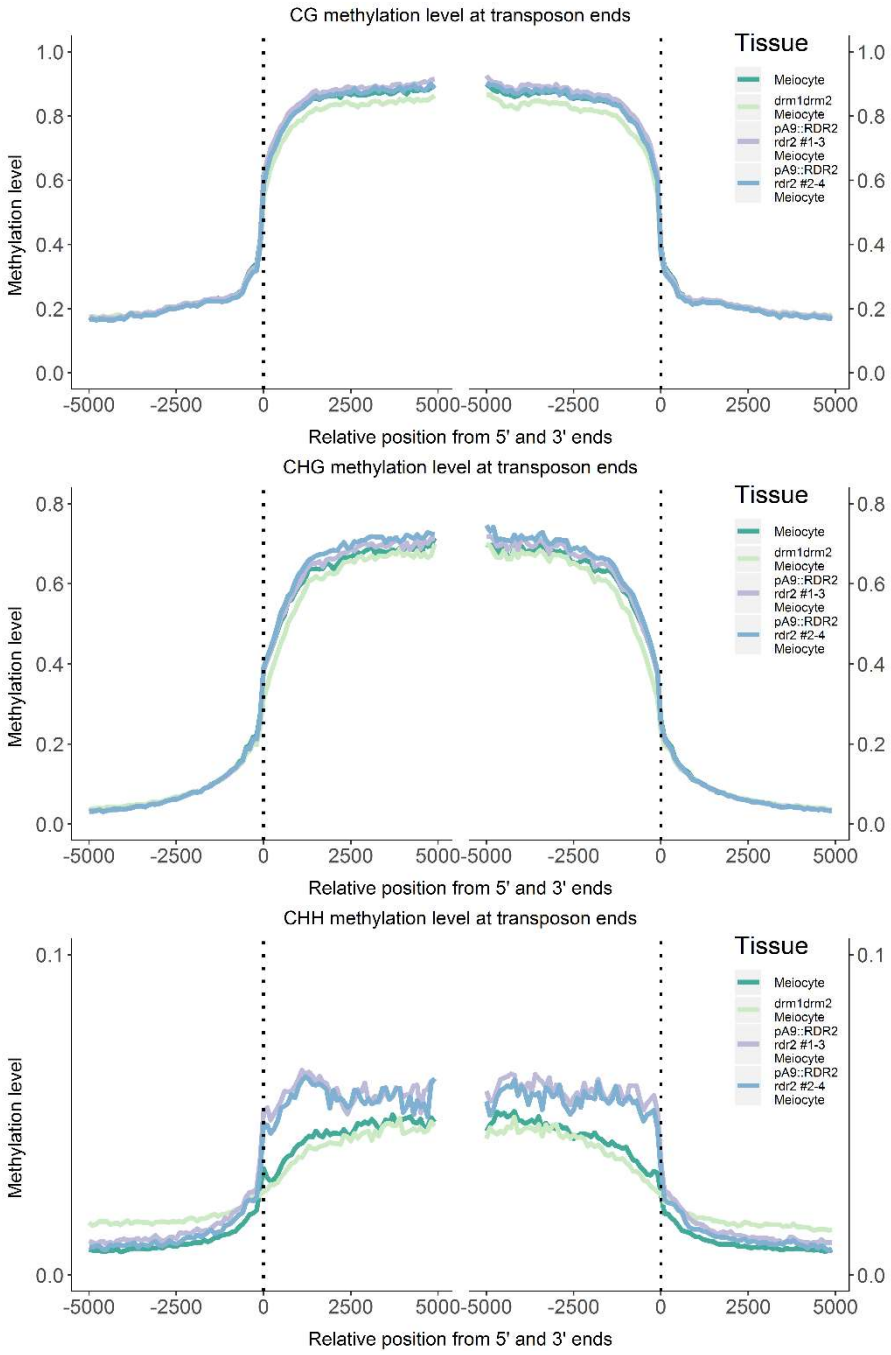


Figure 4-19: Transposon ends analyses of *pA9::RDR2 rdr2* meiocytes (#1-3, #2-4) with WT, and *drm1;drm2* meiocytes. Methylation in the CG and CHG contexts recovers to WT levels or higher in all *pA9::RDR2 rdr2* lines. In the CHH context, methylation at transposon edges is higher than WT levels, showing a clear role of RdDM at these sites. This remains above WT levels over long heterochromatic transposons in *pA9::RDR2 rdr2* meiocytes, suggesting an enhanced role for RdDM.

To understand the contribution of tapetum sRNAs to sexual lineage specific methylation the methylation at SLH and SLMs was investigated in *pA9::RDR2 rdr2* meiocytes and sperm. In all sequence contexts at SLHs, methylation is significantly higher in *pA9::RDR2 rdr2* lines than RdDM-mutant meiocytes or sperm ($p < 0.001$, Conover-Iman test, Fig. 4-20). SLM methylation also shows a pronounced recovery in all sequence contexts, compared to RdDM-mutant lines ($p < 0.001$, Conover-Iman test) (Fig. 4-20). Consistently, the tapetum shows hypermethylation at SLHs but not at SLMs, which are methylated by mismatched sRNAs in the meiocytes (Fig. 4-15). The tapetum can therefore drive the methylation of SLHs and SLMs in the meiocytes by delivery of SLH-derived sRNAs. At both SLHs and SLMs, *pA9::RDR2 rdr2* meiocytes show significantly higher CHH methylation than WT meiocytes ($p < 0.001$, Conover-Iman test) (Fig. 4-20). However, in *pA9::RDR2 rdr2* sperm, CHH methylation at SLHs and SLMs is lower than WT ($p < 0.001$, Conover-Iman test) (Fig. 4-20). This suggests that there is a loss of methylation at these loci through pollen development as the sexual lineage lacks a functional RDR2 in the *rdr2* mutant background.

These data together support the hypothesis that the tapetum can be a source of sRNAs to drive DNA methylation in the meiocytes. Recovery of methylation at transposons and sexual lineage hypermethylated sites in the meiocytes shows sRNAs are transferred before meiosis, and the methylation in sperm suggests that tapetum-derived sRNAs remain active in the sexual lineage.

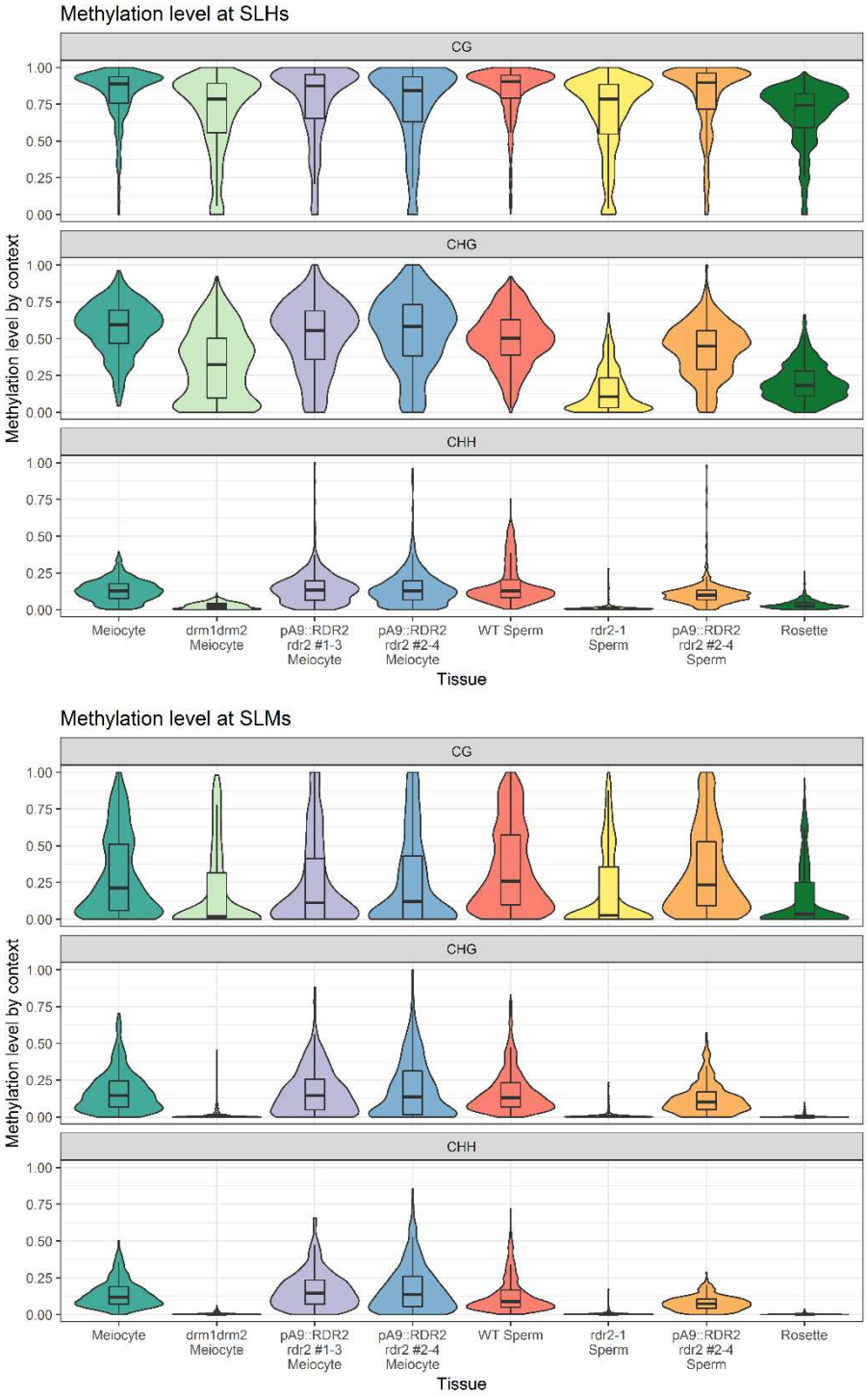


Figure 4-20: Methylation level by sequence context at SLHs and SLMs in *pA9::RDR2 rdr2* meiocytes and sperm, as well as WT meiocyte and sperm, *drm1;drm2* meiocyte, and *rdr2* sperm. Methylation at both SLHs and SLMs shows a strong recovery from mutant levels in *pA9::RDR2 rdr2* sperm and meiocytes. *pA9::RDR2 rdr2* meiocytes show significantly higher CHH methylation than WT, while *pA9::RDR2 rdr2* sperm methylation is significantly lower, suggesting a loss of methylation through sexual lineage development.

4.8 Discussion

These results reveal both specific and shared sites of hypermethylation in the tapetum, as well as sRNA movement from the tapetum to the meiocytes. This work has provided insight into tapetal and sexual lineage DNA methylation dynamics and has also allowed me to propose interesting hypotheses of tapetum function. These hypotheses, however, require further testing.

4.8.1 A novel crosstalk between RDR2- and RDR6-RdDM

Tapetum-specific hypermethylated DMRs depend on both RDR2 and RDR6 for their methylation, showing a role of canonical and non-canonical RdDM pathways in the tapetum. Many DMRs completely lose methylation in both *rdr2* and *rdr6* mutant tapetum, suggesting that the pathways do not act redundantly (Figs. 4-9 & 4-10). Given the known role of RDR6 in establishing DNA methylation via 21-22 nt sRNAs [145, 148] it is possible that RDR6 acts earlier in development to establish hypermethylation, RDR2 can then maintain this methylation through tapetal development. It is possible that RDR6 functions after RDR2, though this is unlikely as DNA methylation inhibits Pol II transcription and so would prevent RDR6 function [145].

The potential for these pathways to be acting sequentially to establish and maintain DNA methylation in a single lineage over developmental time is an exciting hypothesis. The ability to follow establishment and maintenance of methylation through a single cell-type could lead the tapetum to be an interesting model to study the interplay between RdDM pathways. The exact timing of RDR6 and RDR2 activity through tapetum development could be investigated with fluorescent reporter lines. RDR6 has been suggested to function in reproductive tissue precursor cells to silence long centromeric transposons [148]. This function may well lead to the establishment of *de novo* methylation in tapetum precursor cells. While tapetum-specific hypermethylated DMRs show interesting dynamics between RDR2 and RDR6, they have little effect on gene expression and their function remains to be explored.

4.8.2 Evidence of expanded RdDM activity in the tapetum

In the vegetative cell RdDM activity expands into previously heterochromatic loci [166, 178]. In the columella cells also, a loss of heterochromatin is proposed to allow the expansion of RdDM, and so increase the pool of 24 nt sRNAs available to direct DNA methylation [151]. In the tapetum and meiocytes there is possible evidence of an expansion of RdDM activity into previously heterochromatic loci. From transposon ends analysis, CHG methylation along the length of transposons is lower in the *rdr2* mutant than WT (Fig. 4-4), differing from other cell-types where only transposon edges are affected (Fig. 4-1).

In RDR2 complementation experiments, *pA9::RDR2 rdr2* meiocytes, show higher levels of methylation across the full length of transposons (Fig. 4-19), differing from *pA9::RDR2 rdr2* sperm where RdDM-activity is only present at transposon edges (Fig. 4-18). This would suggest that in the meiocytes, Pol V can access previously-heterochromatic transposons to bring about DNA methylation via tapetum-derived 24 nt sRNAs [163]. This fits with the open chromatin environment known to exist in meiocytes [163]. RdDM-induced methylation at long transposons in *pA9::RDR2 rdr2* meiocytes, suggests there is a concomitant reduction of heterochromatin in the tapetum, allowing Pol IV to produce sRNAs from long transposons. This expansion of RdDM-targeting may act to provide an increased pool of sRNAs to maintain methylation at evolving transposon sequences, reflecting the targeting of transposons by piRNAs in animal systems [293, 294]. Alternatively, it may be that 24 nt sRNAs, derived from canonical RdDM-targets, are acting with mismatches to methylate long transposons in *pA9::RDR2 rdr2* meiocytes (Fig. 4-17), or that activity of the CMT methyltransferases is altered in *pA9::RDR2 rdr2* meiocytes [121]. This methylation could be established by RdDM, without a loss of heterochromatin in the tapetum.

While the lack of CMT2 expression in the tapetum suggests that heterochromatin may be reduced, this assertion requires direct testing. Chromatin-immunoprecipitation sequencing experiments could be performed on FACS-sorted tapetal cells to identify the distribution of heterochromatin-associated histone marks [111]. Assay for

transposase accessible chromatin with high-throughput sequencing (ATAC-seq), could also be performed to discover accessible chromatin sites in the tapetum, and show whether there is truly a reduction of heterochromatin [295]. With meiocyte sRNA-seq data, sRNAs mapping to long transposons that show an increase in methylation in *pA9::RDR2 rdr2* meiocytes could be investigated. This would suggest whether RdDM-induced methylation of long transposons was due to an increased pool of 24 nt sRNAs from the tapetum, or via sRNAs acting with mismatches.

4.8.3 The tapetum can provide sRNAs to meiocytes, and thereby affect germline methylation

I have shown that the tapetum possesses hypermethylation at SLHs, which have been shown to be hypermethylated in the sexual lineage [9]. However, the tapetum does not possess methylation at SLMs. This suggests that the mechanism by which DNA methylation is established differs between SLHs and SLMs.

In *pA9::RDR2 rdr2* tapetum, SLHs can produce 24 nt sRNAs that act non-cell autonomously and direct DNA methylation in meiocytes. SLM methylation in meiocytes is also restored upon tapetal expression of RDR2 despite these sites being unmethylated in the tapetum. Evidence from meiocyte sRNA-sequencing suggests that SLH-derived sRNAs can act with mismatches to bring about methylation at SLMs that have high sequence similarity. The lack of detectable sRNAs deriving from SLMs in meiocytes, despite their strong methylation, suggests that Pol IV and Pol V are decoupled in the sexual lineage and that the tapetum is likely to be the major source of meiocyte 24 nt sRNAs. These data together have allowed us to propose a mechanism by which tapetal sRNAs can bring about alternative DNA methylation states in the sexual lineage (Fig. 4-21). SLH-derived sRNAs can be provided by the tapetum to bring about methylation at both SLHs and SLMs in the meiocytes.

SLM methylation being brought about by mismatched sRNAs reflects the situation in the *Caenorhabditis elegans* germline, where piRNA binding with mismatches is permitted in the germline [293, 294, 296]. This may be an evolutionary adaptation to

keep novel transposon insertions and mutations epigenetically silenced and may reflect the greater need to control non-self-elements in germlines than in soma [9, 297].

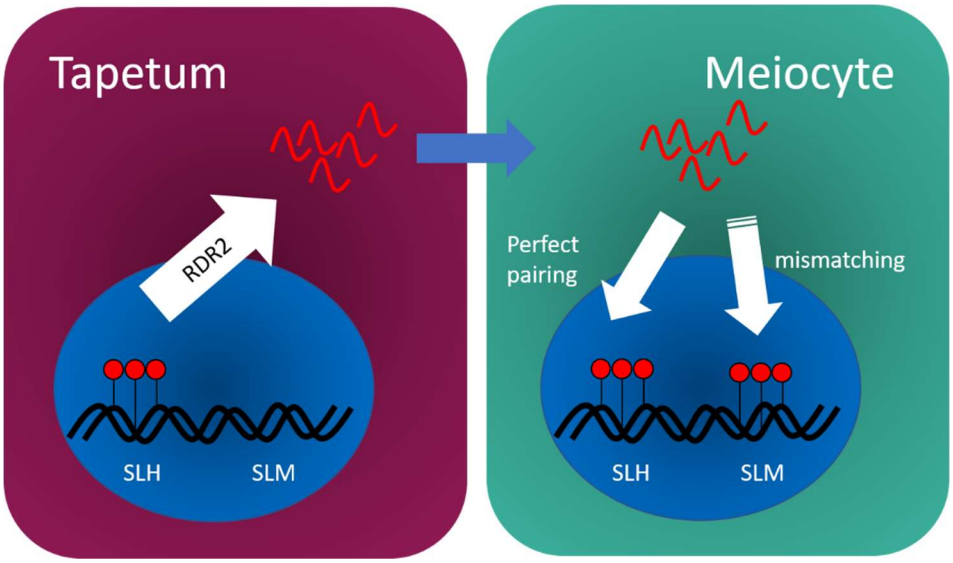


Figure 4-21: Proposed model of the origin of DNA methylation at SLMs. The tapetum possesses strong methylation at SLHs which depend on RDR2 for methylation. RDR2 can produce 24 nt sRNAs from these loci which can act non-cell-autonomously to bring about methylation in the meiocytes. SLHs can be methylated through perfect pairing of sRNAs. SLMs share sequence similarity with SLHs and so can be methylated by SLH-derived sRNAs acting with mismatches. What factor allows mismatches in the meiocytes but not the tapetum is unknown.

Methylation induced by *pA9::RDR2 rdr2* is maintained into the sperm cells, the true germline, and as such may represent a novel route for soma-germline crosstalk. The orthodox view that information can only pass from the germline to the soma is being challenged by further studies of epigenetic inheritance, and while transgenerational inheritance of epigenetic marks is exceptional, it is entirely plausible that tapetal small RNAs could affect the inheritance of such marks via the sperm cells. Methylation at RdDM targets can be maintained in the sperm of *pA9::RDR2 rdr2* plants, which would suggest that sRNAs from the tapetum remain active across meiotic and mitotic cell divisions, which would otherwise dilute DNA methylation, particularly in the non-symmetric CHH context.

4.9 Summary

This chapter has revealed novel DNA methylation in the tapetum as well as methylation patterns shared with the sexual lineage. Tapetum-specific methylation shows a novel dependence on both RDR2 and RDR6, while methylation at SLHs depends solely on RDR2. RdDM loci in the tapetum can provide small RNAs to direct methylation in meiocytes, as shown when RDR2 was expressed in the tapetum in an *rdr2* mutant. Recovery of methylation in the sexual lineage supports a preference of the DRM methyltransferases for CHH sites. Methylation was restored above WT levels across the full length of transposons in *pA9::RDR2 rdr2* meiocytes, suggesting a potential expansion of RdDM activity. Tapetum-derived 24 nt sRNAs can re-establish methylation at SLHs in the sexual lineage, but also at SLMs which are unmethylated in the tapetum. This methylation of SLMs in the *pA9::RDR2 rdr2* sexual lineage is supported by evidence from sRNA-seq which suggests that SLM methylation can be established by mismatched SLH-derived sRNAs. Together this has allowed us to propose a model where the tapetum can provide sRNAs to direct methylation at both SLHs and SLMs in the sexual lineage.

4.10 Materials and Methods

4.10.1 Meiocyte and sperm extraction

Plants were grown as described previously (2.11.1) and meiocyte extraction was performed as per [9]. Sperm nuclei were extracted by FACS using a BD FACS-Melody sorter (BD, Reading UK) using a sperm isolation protocol from [152].

4.10.2 Library production and sequencing

Tapetal cells were isolated as previously (2.11.3). Tapetum bisulphite sequencing libraries were created using the single-cell library method described in [255]. Libraries were created from pools of 400 FACS-sorted cells and sequenced with 100 bp single end reads on a Hiseq 2500 (Illumina).

Bisulphite sequencing libraries from meiocytes and sperm were created using the Ovation Ultra-low methyl-seq (Nugen, 0336) and EpiTect Fast Bisulfite Conversion (Qiagen, 59802) as per [9]. Libraries were sequenced with 100 bp reads on a Hiseq 2500 (Illumina) or with 75bp reads on a Nextseq 500 (Illumina).

RNA-seq libraries from *rdr2* tapetal cells were created using the single-cell system from [188]. Libraries were created from pools of 500 FACS-sorted cells. Libraries were sequenced with 38bp paired end reads on a Nextseq 500 (Illumina). A full list of libraries created is available in Appendix table A1.

4.10.3 Bisulphite-sequencing mapping and data sources

Bisulphite-sequencing data was mapped to the TAIR10 *Arabidopsis thaliana* genome using one of two methods. All tapetum, *pA9::RDR2 rdr2* meiocytes and sperm, and microspore data were analysed using the miniature-sniffle-mapper (MS) developed by Martin Vickers (https://github.com/martinjvickers/miniature-sniffle-mapper). MS is based on Bismark [298], but with the additional ability to map ambiguous reads.

The remaining data was analysed using the in-house script BS-sequel as in [152]. This includes; meiocytes, sperm, vegetative cells, *drm1;drm2* meiocytes, *rdr2* sperm, rosette leaf, cauline leaf, root, and seedling data. WT and *drm1;drm2* meiocytes data are from [9], WT sperm, *rdr2* sperm, and WT vegetative cell data are from [166], microspore data are from [154], cauline leaf from [299], rosette leaf from [127], root and seedling data from [118]

Both methods map ambiguous reads and produce comparable results, but only MS can be used for non-directional sequencing libraries (as produced by [255]).

4.10.4 Methylation analyses

Transposon ends analyses was performed using scripts from [152] and the TAIR10 transposon annotation. Methylation level across annotations was calculated from the single cytosine level through windowing by annotation. Methylation across an

annotation was calculated as the total sequenced Cs divided by the sum of the total Cs and Ts across the annotation. Correlation in 50 bp windows was calculated between biological replicates of WT, *rdr2,* and *rdr6* tapetum using Pearson's product-moment correlation. Average correlations are presented with standard deviations. Control DMR lists were created as in [9].

For sexual lineage average methylation meiocyte, microspore, and sperm data were combined at the single cytosine level. Methylation across annotations were then calculated as above from the combined numbers of Cs and Ts.

4.10.5 Methylation recovery analysis

The percentage methylation recovery in pA9::RDR2 line relative to WT was calculated by first selecting 50 bp windows that are well methylated in the rosette leaf. Windows with CG methylation above 0.7, CHG above 0.45, or CHH above 0.15 were selected. This yielded 398790 windows for CG, 163630 for CHG, and 202373 for CHH. Methylation recovery was investigated independently for each set of windows. The methylation score was calculated for each context for their respective windows. Average methylation across all windows was calculated for each context, and for each genotype. Methylation scores are relative to WT for each tissue.

4.10.6 Differential expression

Differentially expressed genes were found using TopHat and Cufflinks [234], aligning to the TAIR10 gene annotation.

4.10.7 DMR analysis

Differentially methylated regions were found using methpipe [300]. DNA methylation data from three independent biological replicates of tapetum BS-seq data was compared to published data from root, seedling, rosette leaf, and cauline leaf (section 4.10.3). Methylation was compared at the single cytosine level to a significance of $p < 0.05$ for each sequence context independently and DMRs expanded

in steps of 200 bp [300]. DMRs were combined if CHH-DMRs overlapped CG- or CHG-DMRs. Combined DMRs below 20 bp were removed and were merged if they were within 50 bp of another merged DMR. Merged DMRs were kept if total cytosine methylation (C_{methyl}) between Tapetum and soma was significantly different ($p < 0.001$, Fisher's exact test). Significant DMRs were also required to have higher methylation in the tapetum in the CHH context and either CG or CHG context. To find tapetum RDR2-targets methylation in WT and *rdr2* tapetum were compared at significantly hypermethylated DMRs and those with a significant difference were retained ($p < 0.001$, Fisher's exact test). DMRs that overlapped SLM/SLHs were removed as mentioned in results.

Hypomethylated sites were found by the same method but looking for significantly lower methylation in the tapetum. Final DMRs were kept if they were significantly different between the tapetum and somatic tissues (Fisher's exact test, $p < 0.001$) and lower methylation in the tapetum in the CHH context, as well as either CHG or CG contexts.

Scripts for DMR analysis can be found at (https://github.com/Baldridge37/methpipe)

For SLMs and SLHs, hypermethylation in the tapetum was defined as DNA methylation significantly higher in the CHH context, as well as the CHG or CG contexts ($p < 0.001$, Fisher's exact test, 4.10.10).

4.10.8 Small RNA mapping

Small RNA mapping was performed using Bowtie [301]. In the first analysis sRNAs were mapped with up to 3 mismatches and the best match kept. In the second they were again allowed to map with up to 3 mismatches, but all possible hits were kept, and the mapping at loci weighted by the number of mapped sites for a read. In the first analysis, almost all sRNAs in the output map with 0 mismatches.

4.10.9 Immunostaining

Immunostaining was performed using the method available at https://langdalelab.files.wordpress.com/2015/07/immunolocalization_ap.pdf without edits. The primary antibody was a 1:100 dilution of α-FLAG MONOCLONAL ANTI-FLAG (Sigma-Aldrich, F1804). The secondary antibody was a 1:1000 dilution Alexa-Fluora 555 (Thermo Fisher, A21424).

4.10.10 Statistics for comparison of methylation levels

For comparisons of methylation levels at specific annotations a Kruskal-Wallis test was performed as methylation data was not normally distributed. To find pairwise differences between samples *post-hoc* analysis was performed using the Conover-Iman test and implementing the Holm correction to control false discovery rate ($p < 0.05$). After Holm correction significance cut-off was $p < 0.025$.

For Fisher's exact test the ratio of C and T reads across an annotation were compared, and 0.001 was taken as the *p*-value cut-off for significance.

4.10.11 Creation of *pA9::RDR2:FLAG rdr2* lines

pA9 was amplified with pHG001 and pBA008 (Primers, section 6) to give an attB1 site and a 24 bp overlap with the TSS of *RDR2*. The genomic sequence of *RDR2* was amplified with pBA009 and pBA010 to give a 24 bp overlap with *pA9* and an attB2 site, while skipping the stop codon (primers, section 6). Fragments were combined through overlapping PCR and cloned into the entry vector pDONR207. Correct clones were then incorporated into the pGWB10 destination vector to give the full *pA9::RDR2:FLAG* construct [302]. *Arabidopsis thaliana rdr2-1* plants were then transformed by floral dip with *Agrobacterium tumefaciens GV3101* [233]. Transformed seeds were surface sterilised and sown onto full strength Murashige & Skoog media (no glucose) containing hygromycin (25 ng/mL). Seeds were stratified for 2 days at 4 °C before being grown for one week with 16-h light 8-h dark at 20 ^{0}C. Positive transformants were transferred to soil. Homozygous T3 plants were analysed after

transgene inheritance was assayed by segregation of Hygromycin resistance. All lines analysed possessed single inserts (from segregation) and showed no *RDR2* expression in leaves.

Chapter 5 General Discussion

The broad aim of this thesis has been to investigate tapetum's function in supporting germline development through the application of sequencing technologies. While each chapter has explored a different aspect of tapetum biology, and employed different technical approaches, there are links between these bodies of work which provide greater insight into tapetum biology. The knowledge gained from the tapetum transcriptome, detailed in Chapter 2, is expanded on by the temporal dynamics inferred in Chapter 3. The inference of stage-specific expression of genes also adds depth and context to the knowledge gained about known and novel tapetal functions in all chapters. I will discuss insights from studying both transposon expression and DNA methylation, and the implications for tapetal function. I will elaborate on the model of tapetal small RNA movement proposed in Chapter 4, bringing in further evidence from other chapters and putting this model in context of the literature. The thesis concludes with a reflection on the advances made in this work, and an outlook on its relevance to the study of sexual reproduction and epigenetics.

5.1 The value of high quality tapetum transcriptomes

While the tapetum is an important tissue in the study of male germline development it is only recently that whole genome RNA-sequencing data has become available. While collecting tapetal cells for this thesis required the time-consuming optimisation of protoplasting and fluorescence-activated cell sorting protocols, I believe this approach was worthwhile. I have produced high-purity bulk and single-cell RNA-sequencing libraries of mRNA, as well as libraries from total RNA and bisulphite-treated DNA. This has allowed a more detailed investigation of tapetum biology than was previously possible.

An advantage of examining gene expression across the whole genome is the ability to find previously unknown genes. From this data we can look in greater detail at

those genes previously described and find new genes to explore the full suite of tapetal gene expression. In Chapters 2 and 3 I explored the tapetal transcriptome at both a bulk and single-cell level. This has allowed me to expand knowledge of the genes functioning in the tapetum generally and at specific developmental timepoints.

5.2 The identification of novel genes expressed in the tapetum

In this thesis I have identified several tapetum-enriched and tapetum-specific genes which have been implicated or shown to be involved in tapetum and pollen development. Plants with a T-DNA insertion in the *DF1* gene show defects in tapetum and pollen morphology, with a possible reduction in pollen yield. The downstream targets of DF1 are unclear at this point. In a recent study DF1 was found to negatively regulate cell wall biosynthetic processes, including pectin biosynthesis [202]. This suggests that DF1 regulates tapetal and pollen wall development [42], which may underlie the pollen and tapetal defects seen in *df1*. The tapetum-enriched transcription factor MYB52 has also been implicated in pectin metabolism in the seed coat [199]. This may suggest previously unknown similarities between seed coat mucilage and pollen coat/wall biosynthesis which may represent a new avenue of research.

In pseudotime *MYB52* shows a late expression profile, but *DF1* could not be detected in single-cell data. How these factors could interact to affect pectin metabolism in the tapetum remains to be explored. Indeed, MYB52 could also function with the C2H2 zinc-finger protein AT1G02040 and the helix-turn-helix protein AT1G18960. Both *AT1G02040* and *AT1G18960* show late pseudotime expression profiles. Pollen phenotypes and double mutants could be explored further to identify possible roles in pollen wall or pollen coat metabolism.

From the tapetum transcriptome I have been able to identify genes which are expressed only in the tapetum when compared to 11 other tissues. LBD2 is a transcription factor which I investigated in greater detail and found to be required

for normal tapetal and pollen development. *lbd2* plants were found to possess vacuolated and multi-layered tapetal cells, and a higher rate of misshapen pollen (Chapter 2). These misshapen pollen grains showed defects in pollen wall structure. From pseudotime analysis *LBD2* shows a peak of expression late in tapetal development but is highly variable between individual single-cells (Chapter 3). While average expression is low, some cells show very high levels of expression, suggesting a strong but short peak of expression. *LBD2* also shows a remarkable level of co-expression with *SND3/NAC10*, which possibly suggests the genes are coregulated and act together. *SND3/NAC10* has been shown to be a positive regulator of secondary cell wall synthesis [205], further suggesting *LBD2* regulates pollen wall biosynthesis. This highlights the value in exploring the tapetum transcriptome at both a bulk and single-cell level as differences within and between cell types.

While I have explored the function of several tapetum-expressed transcription factors in detail, it is beyond the scope this thesis to investigate all those proposed. I hope that both the bulk and single-cell RNA-sequencing data will prove a great asset to other researchers exploring tapetum function and male fertility. The techniques developed in this thesis will facilitate further research on these genes, for example, by allowing the isolation of high-purity cells from mutant plants. Through this approach a fuller picture of tapetum transcriptional dynamics and gene regulation can be built.

5.3 Implications of tapetal transposon expression

Throughout the life cycles of plants there are periods of transposon activation and expression, particularly during reproductive development [105]. Whether this developmentally-timed expression is a strategy to produce small RNAs, serves a developmental function, or is just the deleterious expression of selfish elements is still debated. One such period of transposon expression is in the male meiocytes where estimates range from 5% to 32.5% of all transposons expressed [106, 107]. In pollen, the vegetative cell undergoes a relaxation of transposon silencing to produce sRNAs

that reinforce silencing in the sperm cells [152, 154]. Whether transposons are also expressed in the tapetum has remained unclear. In drosophila, it has been shown that nurse cell-derived transposons can move to the germline where they preferentially insert, rather than in the nurse cell genome [303]. Nurse cell expression of transposons may therefore not be a risk-free strategy to tackle germline transposon expression. While there has been some evidence of transposon expression in the tapetum of *Zea mays* [304], a genome-wide understanding of transposon expression has been lacking.

From total RNA-seq libraries (Chapter 2) I have identified transposons expressed in the tapetum. When compared to meiocytes, pollen, and rosette leaf, the tapetum was found to be more like leaves than the sexual lineage cell types, in both the number and type of transposons expressed. Of those transposons that are well-expressed in the tapetum, very few are tapetum-specific, and most are expressed to a similar level in other tissues. This suggests that the tapetum does not participate in the meiocyte relaxation of transposon silencing. It is possible that transposon expression is restricted to specific stages, and that investigating expression in bulk RNA-sequencing of mixed stages has obscured the peak of transposon expression. The early pseudotime expression of genes functioning in RdDM and chromatin assembly (Chapter 3) suggests that chromatin changes in the tapetum occur at specific stages. It is feasible that these chromatin changes could be a cause or consequence of transposon expression. Transposon expression cannot be investigated at the single cell level as mRNA is pulled-down by their polyA tails, and most transposon transcripts are not polyadenylated. Novel transposon insertions could be detected by single-cell genome sequencing, however. Sorting tapetal cells of specific stages for total RNA-sequencing would reveal stage-specific transposon expression, using markers identified in Chapter 3.

Lack of tapetum transposon expression may be surprising given the expression which occurs in the vegetative cell. Firstly, comparisons between the vegetative and tapetal nurse cells may be spurious given the different functions performed by these cells. The vegetative cell undergoes large scale chromatin changes including a loss of

centromere identity [178]. This is possible as the vegetative cell does not perform any further cell divisions and will not contribute to the next generation. While the tapetum is also a short-lived cell type it will undergo endoreplication and nuclear division. Endoreplication occurs once the tapetum is isolated from the meiocytes by a callose cell wall (after putative sRNA transfer). The tapetum must also support the high level of metabolic activity required for the rapid production of pollen wall and coat components. Chromatin changes on the scale seen in the vegetative cell may therefore not be compatible with other functions of the tapetum. Secondly, it is possible that the tapetum does support the meiocyte in silencing transposons, but that the large-scale expression of transposons is not required for this function.

5.4 The tapetum as a site of DNA methylation reprogramming

In Chapter 4 I explored DNA hypermethylation in the tapetum. When the tapetum was compared to other somatic tissues, I found tapetum-specific hypermethylated DMRs which are dependent on both RDR2 and RDR6 for methylation. It was found that this was not due to the presence of separate RDR2- and RDR6-dependent clusters, but the near-equal dependence on both RNA-dependent RNA polymerases. The question of how two alternative sources of small RNAs can both be required for methylation at these sites is a pertinent one. If the two pathways were acting simultaneously, then one would expect there to be redundancy in the maintenance of methylation. As such, one would predict that, in both the *rdr2* and *rdr6* mutants, methylation would be maintained by the other pathway. Instead we see a near equal loss of methylation in both mutants toward the low level of methylation seen in WT rosette leaf.

To explain this, I propose that RDR2 and RDR6 are acting in series through the development of the tapetum to establish and maintain methylation at these loci. Other research has shown that RDR6-derived sRNAs can bridge post-transcriptional and transcriptional gene silencing through ARGONAUTE 6 (AGO6) [145, 148]. This

must occur prior to the stage of the tapetal cells collected as *AGO6* is undetected in bulk RNA-sequencing data. The RNA-directed DNA methylation pathway can then maintain methylation after establishment by RDR6-AGO6. It is possible that tapetum-specific hypermethylated DMRs are transcribed by Pol II and acted on by RDR6, but these loci do not significantly overlap expressed genes or transposons. It may be that they are expressed in tapetal precursor cells or that other loci are the sRNA source. Producing tapetal sRNA-seq libraries would further our understanding of the contributions of these pathways. Single-cell mRNA and bisulphite-sequencing [263] would allow the tracking of DNA methylation through development, where single cells can be positioned along a developmental axis through pseudotime analysis.

5.5 The tapetum as a source of small RNAs

In anthers of grass species, such as rice and maize, the tapetum is a source of 24 nt phased small interfering RNAs (PhasiRNAs) to the developing male sexual lineage [156, 170, 305]. While there is no evidence to suggest that *Arabidopsis* possesses reproductive phasiRNAs, experiment conducted in Chapter 4 suggest that the tapetum can be a source of RDR2-derived 24 nt sRNAs. When RDR2 is expressed in the tapetum in an *rdr2* mutant, methylation at RdDM-target loci is restored in both meiocytes and sperm (Chapter 4). Early in anther development, the tapetum and meiocytes are connected via plasmodesmata [15]. While meiosis progresses, meiocytes are isolated from the rest of the plant by a callose cell wall. Once the callose cell wall is broken down, the tapetum quickly produces the sporopollenin pollen wall. One may therefore expect that the early connections between tapetum and meiocytes are the only opportunity for sRNA transfer, but a recent paper suggests that sRNAs can also be transferred in extracellular vesicles [306]. While vesicle sRNA transfer was in response to a pathogen, it is possible that this mechanism is also employed in normal development. To test this, RDR2 complementation experiments could be repeated, but under a late-stage tapetum-specific promoter, which analyses in Chapters 2 and 3 have identified. If methylation is restored in plants expressing

RDR2 in the tapetum after the isolation of the meiocytes, then it may suggest a non-plasmodesmal route of sRNA transfer.

While I have shown that tapetal expression of RDR2 is sufficient to drive meiocyte methylation (at least at a high level of expression) it is unknown if tapetal expression of RDR2 is necessary in WT plants. Meiocytes lack 24 nt sRNAs derived from SLMs, despite SLMs being hypermethylated, suggesting that the Pol IV and Pol V pathways are decoupled, and meiocytes depend on another source of sRNAs. To test this fully would require the generation of plants lacking RDR2 in the tapetum but with a functional copy of RDR2 in the meiocytes. Similar complementation experiments could be performed, as for *pA9::RDR2 rdr2*, but with a meiocyte-specific promoter driving RDR2 expression in an *rdr2* mutant.

In the complementation experiments performed with *pA9::RDR2 rdr2* plants, the high expression of RDR2 from the *A9* promoter may lead to artefacts. Weaker tapetum-specific promoters (Chapters 2 and 3) could be used to test this. A more naturalistic approach may be to complement an *rdr2* mutant with a functional *RDR2* under control of the native promoter. Inserting *loxP* sites into introns along with cell-type specific expression of the Cre recombinase would allow the removal of exons between the *loxP* sites, and so create a tissue-specific mutant [307]. Reciprocal experiments with tapetum- and meiocyte-specific mutants could be performed for direct comparison. Experiments could also be performed to create a cell type-specific rescue of *RDR2* under control of the native promoter.

An unexpected result from tapetum RDR2-complementation experiments was the re-methylation of SLMs in the meiocytes. SLMs lack methylation in the tapetum, suggesting that 24 nt sRNAs are not sourced from these loci. Evidence from meiocyte sRNA-seq has suggested that sRNAs derived from SLHs can act with mismatches to methylate SLMs. We therefore propose a model to explain these data (Fig. 4-21). RdDM is recruited to SLHs in the tapetum, where hypermethylation occurs and abundant 24 nt small RNAs are produced. These sRNAs can then move to the meiocytes. Methylation occurs at SLHs in the meiocytes through RdDM, but

mismatches are permitted, and methylation also occurs at SLMs *de novo*. Canonical RdDM then takes over maintenance of SLM methylation through sexual lineage development and SLM-derived sRNAs are present in the mature pollen. This model can elegantly explain results seen from RDR2 complementation experiments and sRNA-seq and also provides testable hypotheses of sexual lineage DNA methylation.

This model implies that some factor, or factors, differ(s) between the tapetum and meiocytes allowing mismatches in the latter but not the former. High purity transcriptomes of both the tapetum and meiocytes are invaluable in deciphering differences in gene expression which could explain the divergent DNA methylation patterns. Prime candidates for this factor are the sRNA-binding ARGONAUTE proteins. From RNA-seq data the only AGO gene significantly more highly expressed in the meiocytes than the tapetum is *AGO6*. If AGO6 could be responsible for 24 nt mismatching is unclear, as evidence suggests AGO6 binds 21-22 nt sRNAs [148]. Of other RdDM components, the *DOMAINS REARRANGED METHYLTRANSFERASE* genes also show differential expression. *DRM2* is significantly more highly expressed in the tapetum, while near-absent in the meiocytes, with the opposite true for *DRM1*. Whether methyltransferases could affect sRNA base-pairing is unknown at this time.

This model also poses the question of how SLH methylation could be established in the tapetum. A recent paper has shown that the CLASSY family of chromatin remodellers recruit Pol IV to specific loci to generate 24 nt sRNAs [130]. Of the four *CLASSY* genes two are differentially expressed between the tapetum and meiocytes. *CLSY4* is significantly higher in the meiocytes and *CLSY3* is significantly higher in the tapetum, while both genes are very lowly expressed in leaves. *CLSY3* and *CLSY4* were shown to synergistically regulate 24 nt sRNA production at pericentromeric loci [130]. The differential expression of *CLSY3* and *CLSY4* could explain how Pol-IV is recruited to SLHs in the tapetum, and so produce sRNAs to bring about DNA methylation in the meiocytes [130].

The model relies on the ability of SLH-derived sRNAs to direct DNA methylation with mismatches at SLMs. In *Caenorhabditis elegans,* piRNAs have been shown to bind

targets with mismatches, theoretically targeting every transcript [293, 294]. This mismatching has been proposed to allow targeting of rapidly evolving transposon sequences. While every transcript can be targeted, germline gene expression is protected by a specific gene-licensing argonaute [308], suggesting that piRNA silencing is the default state in the germline. The self-incompatibility system of *Arabidopsis halleri* relies on genetic dominance relationships facilitated by sRNA-mediated silencing [309]. sRNAs are able to repress the expression of recessive alleles through binding with up to three mismatches [310]. This suggests that sRNA mismatching is more widespread than previously thought and also functions outside of the sexual lineage. It may be that DNA methylation can be brought about by mismatched sRNAs in all tissues, but in the meiocytes Pol IV and Pol V decoupling allow us to see mismatching before sRNAs are produced from methylated loci.

Small RNAs have been proposed to move between cells to reinforce silencing/DNA methylation in many plant systems, but the exact mechanism by which sRNAs are produced, and their mode of action, varies between systems. I have already discussed reproductive phasiRNAs, but this system is unique to grasses. In the vegetative cell of pollen and central cell of the ovule, DNA demethylation by DME leads to a reinforcing of silencing in the germlines through 21-22 nt sRNAs [109, 152, 154, 179]. In the roots, the columella cells have been proposed as a source of small RNAs to maintain DNA methylation in the stem cells [151]. This is proposed to be achieved through a loss of heterochromatin and an upregulation of RdDM leading to global DNA hypermethylation and sRNA production [151]. Small RNA movement can also be achieved without DNA hypermethylation, and 24 nt sRNAs have been detected moving through the phloem [153]. In fact in *Nicotiana benthamiana*, it was shown that sRNAs are able to move across a graft junction from the roots, to silence genes in the male sexual lineage [311]. This suggests that the tapetum may load sRNAs to the sexual lineage which were produced further away.

These examples highlight the diversity and malleability of sRNA pathways employed by plants as well as the uniqueness of each system in which they are

employed. The *Arabidopsis* tapetum appears to function as a source of small RNAs without global hypermethylation or demethylation, but rather with DNA hypermethylation directed at specific loci. This then leads to hypermethylation at matching and novel loci in sexual lineage cells. This, therefore, represents a unique system of sRNA movement, representing information transfer from the soma to the germline.

5.6 Concluding remarks

The development of the germline represents a crucial window in the life cycle of multicellular organisms, one which has repercussions on an individual and evolutionary timescale. The support of germline development by nurse cells, which for the male germline in plants is the tapetum, is an essential aspect of this process, and for plants allows the development of functional pollen.

This thesis has aimed to gain a deeper understanding of the *Arabidopsis* tapetum through the application of cell sorting and sequencing technologies. While each chapter is distinct in its aims and methods used, the understanding gained from one chapter has fed into others; providing a greater knowledge of tapetum function. Together these data have allowed me to identify novel genes expressed throughout tapetal development, as well as those expressed at specific stages. The tapetum was shown to possess novel DNA methylation patterns, while also sharing sites hypermethylated in the sexual lineage. I have proposed that the tapetum is a source of 24 nt sRNAs to direct DNA methylation in the meiocytes, which represents a novel route of soma-germline crosstalk.

This tool development and descriptive research has provided a wealth of data from which numerous hypotheses of tapetum and germline development have been derived. I hope these data will support further exploration of the tapetum and the male sexual lineage both within the laboratory and in the wider research community.

Primers

Primer name	Sequence	Target
Cloning		
PHG1.pA9 – attB1-F	ggggacaagtttgtacaaaaaagcaggcttcGGATTATAATAATG TGTAGACATTGTAGG	pA9
PHG2.pA9 – attB2-R	ggggaccactttgtacaagaaagctgggtcTCTAATTAGATACTA TATTGTTTGTACTTCTG	pA9
pBA008_pA9-RDR2-overlap	GTTCGTCGTCGTCTCTGACACCATTCTAATTAGATACTAT ATTGTTTG	pA9-RDR2 overlap fragment
pBA009_pA9-overlap-RDR2	CAAACAATATAGTATCTAATTAGAATGGTGTCAGAGAC GACGACGAAC	RDR2-pA9 overlap fragment
pBA010_RDR2-NOSTOP-attB2	ggggaccactttgtacaagaaagctgggtcAATGGATACAAGTCC ACTTG	genomic RDR2 reverse pA9::RDR2

Primer name	Sequence	Target
q-RTPCR		
pBA147	TTGAAAAGTGGAACCGTTCTG	MVD1 AT2G38700.1
pBA148	AGGAGGGAAGCAGTAGAGCAG	MVD1 AT2G38700.1
pBA149	GATTTGTGGCTCTTTCCCTTC	DTX35 AT4G25640.2
pBA150	TCCAGCCTTCATTACACCAAC	DTX35 AT4G25640.2
pBA151	CTACCGGAAGTGATGAAGCTG	ABCG1 AT2G39350.1
pBA152	AACCCAAAAGCAACTGTGATG	ABCG1 AT2G39350.1
pBA153	TGCAAACCTACATTGGAGGAG	AT2G21430
pBA154	CGGTTTTTCCTTTAACCTTGC	AT2G21430
pBA155	ATCACGTCTACGGGTGCTATG	AMC1 AT1G02170.1
pBA156	CCACTACCACCACCATCATTC	AMC1 AT1G02170.1
pBA157	AGTGGCTCATAACCACATTGC	LAP3 AT3G59530.1
pBA158	AAAAGACAAACCGGTCCAAAG	LAP3 AT3G59530.1
pBA159	CAAAGCCGTGATCTTGAAAAG	FKP1 AT4G11820.2
pBA160	ATGCAGCGTAGAGAGAAGCAG	FKP1 AT4G11820.2
pBA161	TTGAGTTGGAGGGTGAGACTG	NRPB6 AT2G04630.1
pBA162	ATCAATCACCACCGACTTGAC	NRPB6 AT2G04630.1
pBA163	GGAGCAGCAAGATGACTATGC	ACAT2 AT5G48230.2
pBA164	CAACAATGGTTGATGGCCTAC	ACAT2 AT5G48230.2
pBA189	TATCAGCGGGAATGAGAAGG	DF1-mRNA F
pBA190	GGCAAGTCTTGGAATCTTCG	DF1-mRNA R
pBA191	TATCGCCGTATTTCCCAGCC	LBD2-mRNA F
pBA192	AGAGAGCTTGGTACGGTCCT	LBD2-mRNA R
pBA193	AACCTCGGAATGGCACACAT	bZIP18-mRNA F
pBA194	GGAGGGTTTGTGCGAGAAGA	bZIP18-mRNA R
PHG101	TACGCCAGTGGTCGTACAAC	ACT8 F
PHG34	GGTAAGGATCTTCATGAGGTAATCAG	ACT8 R

Primer name	Sequence	Target
Genotyping		
pBA043-LBb1.3	ATTTTGCCGATTTCGGAAC	Genotyping all SALK lines
pBA044_SALK_021494_LP	GGGCATATCACGAGAGTTATTTG	SALK_021494 AT1G02040
pBA073_SALK_021494_RP	CAACACGATGAGGACAACATG	SALK_021494 AT1G02040
pBA046_SALK_043542_LP	TGGGTCAATTGACCATAAAGG	SALK_043542 GT3a
pBA075_SALK_043542_RP	AGCCAAAAGCTCCTTTGTCTC	SALK_043542 GT3a
pBA054_SALK_118697_LP	TTATGAGGTTTAGAGCCAGCG	SALK_118697 NEV
pBA083_SALK_118697_RP	GGAAACCGAAGAAAATGAAGG	SALK_118697 NEV
pBA056_SALK_106258_LP	TGCATGCGCAATAGTATATGC	SALK_106258 DF1 (df1-1)
pBA085_SALK_106258_RP	AAATTTCGATAATTGGCCACC	SALK_106258 DF1 (df1-1)
pBA058_SALK_132285_LP	TTTATCATCATTTTGGTCGCC	SALK_132285 bZIP28
pBA087_SALK_132285_RP	TATCCCCTAACAGGATACGGC	SALK_132285 bZIP28
pBA060_SALK_110712_LP	GAAGCAAATGTGTTTGGATCG	SALK_110712 bZIP18
pBA089_SALK_110712_RP	CGGGTTATCAGACCTAGGAGC	SALK_110712 bZIP18
pBA062_SALK_000287_LP	TGGTTATCGCGATTTCATTTC	SALK_000287 NAC10/SND3
pBA091_SALK_000287_RP	CTCGAGGTTAAAGTTACGCCC	SALK_000287 NAC10/SND3
pBA063_SALK_033970_LP	AGTAACTATGGAAGGGCCGAG	SALK_033970 MYB124
pBA092_SALK_033970_RP	AAAAACTTCCAGGCCGATATG	SALK_033970 MYB124
pBA064_SALK_128611_LP	CTCTGATACGGTGATACCAATTG	SALK_128611 AT1G18960
pBA093_SALK_128611_RP	AGATCATGTCACTCACCGTCC	SALK_128611 AT1G18960
pBA067_SALK_049806_LP	AAACACGCTGGAGATGATGAG	SALK_049806 AT1G20670
pBA096_SALK_049806_RP	GTCGTTGTCTTCACCTTCGTC	SALK_049806 AT1G20670
pBA069_SALK_138624_LP	AAAAGGATGTTCATTTGGTGG	SALK_138624 MYB52
pBA098_SALK_138624_RP	TGCAAGTAAATGAGTAATGGTGC	SALK_138624 MYB52
pBA119-SALK_112456_LP	CGAGATTTTGCTTCACAGAGG	SALK_112456 LBD2
pBA121-SALK_112456_RP	AGAGGACCGTACCAAGCTCTC	SALK_112456 LBD2

Glossary

Features

UTR	Untranslated region
TSS	Transcription start site
TES	Transcription end site
T-DNA	Transfer DNA from *Agrobacterium tumefaciens*

Units

bp	nucleotide (base pairs)
FPKM	fragments per kilobase per million reads
kb	kilobases
Methylation level	The number of cytosine reads mapped to a cytosine in the genome, divided by the number of matched cytosines and mismatched thymines. Thymines arise from the bisulphite treatment of unmethylated cytosine, while methylated cytosines are protected from conversion
nt	nucleotides
RPKM	reads per kilobase per million reads
TPM	transcripts per million

Chemicals

BSA	Bovine serum albumin
DTT	Dithiothreitol
EDTA	Ethylenediaminetetraacetic acid
MES	2-(N-morpholino)ethanesulphonic acid
SDS	Sodium dodecyl sulphate
Tris-HCl	2-Amino-2-(hydroxymethyl)-1,3-propanediol hydrochloride

Proteins

A7	ARABIDOPSIS THALIANA ANTHER 7 (AT4G28395)
A9	Tapetum-specific protein A9 (AT5G07230)
ABCG1	ATP-BINDING CASSETTE G1 (AT2G39350)
ABCG26	ATP-BINDING CASSETTE G26 (AT3G13220)
ACAT2	Acetoacetyl-CoA thiolase 2 (AT5G48230)
ACOS5	ACYL-COA SYNTHETASE 5 (AT1G62940)

ACT8	ACTIN 8 (AT1G49240)
AGO4	ARGONAUTE 4 (AT2G27040)
AGO6	ARGONATURE 6 (AT2G32940)
AGO9	ARGONAUTE 9 (AT5G21150)
AMC1/ LOL3	ARABIDOPSIS THALIANA METACASPASE 1 (AT1G02170)
AMS	ABORTED MICROSPORES (AT2G16910)
AT1G02040	C2H2-type zinc finger family protein
AT1G18960	myb-like HTH transcriptional regulator family protein
AT1G20670	DNA-binding bromodomain-containing protein
AT2G21430	Papain family cysteine protease
bZIP18	BASIC LEUCINE-ZIPPER 18 (AT2G40620)
bZIP28	BASIC LEUCINE-ZIPPER 28 (AT3G10800)
CLSY3	CLASSY 3 (AT1G05490)
Chr31	
CLSY4	CLASSY 4 (AT3G24340)
Chr40	
CMT2	CHROMOMETHYLASE 2 (AT4G19020)
CMT3	CHROMOMETHYLASE 3 (AT1G69770)
CYP703A2	CYTOCHROME P450, FAMILY 703, SUBFAMILY A, POLYPEPTIDE 2 (AT1G01280)
CYP704B1	CYTOCHROME P450, FAMILY 704, SUBFAMILY B, POLYPEPTIDE 1 (AT1G69500)
DCL3	DICLER-LIKE 3 (AT3G43920)
DDM1	DECREASE in DNA METHYLATION 1 (AT5G66750)
DF1	Duplicated homeodomain-like superfamily protein, DF1 (AT1G76880)
DME	DEMETER (AT5G04560)
DRM1	DOMAINS REARRANGED METHYLTRANSFERASE 1 (AT5G15380)
DRM2	DOMAINS REARRANGED METHYLTRANSFERASE 1 (AT5G14620)
DTX35	DETOXIFYING EFFLUX CARRIER 35 (AT4G25640)
DYT1	DYSFUNCTIONAL TAPETUM 1 (AT4G21330)
EMS1	EXCESS MICROSPOROCYTES 1 (AT5G07280)
FKP1	FLAKY POLLEN 1 (AT4G11820)
GT3a	Homeodomain-like superfamily protein (AT5G01380)
HEN1	HUA ENHANCER 1 (AT4G20900,AT4G20910)

LAP3	LESS ADHERENT POLLEN 3 (AT3G59530)
LAP5	LESS ADHERENT POLLEN 5 (AT4G34850)
LAP6	LESS ADHERENT POLLEN 6 (AT1G02050)
LBD2	LOB-DOMAIN CONTAINING 2 (AT1G06280)
MET1	METHYLTRANSFERASE 1 (AT5G49160)
MS1	MALE STERILITY 1 (AT5G22260)
MS188/ MYB103	MALE STERILITY 188 (AT5G56110)
MS2/FAR2	MALE STERILITY 2 (AT3G11980)
MVD1	MEVALONATE 5-DIPHOSPHATE DECARBOXYLASE 1 (AT2G38700)
MYB124/ FLP	MYB DOMAIN 124/FOUR LIPS (AT1G14350)
MYB52	MYB DOMAIN PROTEIN 52 (AT1G17950)
NAC010/ SND3	NAC DOMAIN CONTAINING PROTEIN 10/ SECONDARY WALL-ASSOCIATED NAC DOMAIN PROTEIN 3 (AT1G28470)
NEV	NEVERSHED (AT5G54310)
NRPB6B	RNA polymerase Rpb6 (AT2G04630)
RDR2	RNA-DEPENDENT RNA POLYMERASE 2 (AT4G11130)
RDR6	RNA-DEPENDENT RNA POLYMERASE 6 (AT3G49500)
SPL/NZZ	SPOROCYTELESS/NOZZLE (AT4G27330)
TDF1	DEFECTIVE in TAPETAL DEVELOPMENT and FUNCTION1 (AT3G28470)
TPD1	TAPETAL DETERMINANT 1 (AT4G24972)

Appendix

Library	Genotype	Cell/ Tissue	Library type	Reads (M)	Read length (bp)	Single/ Paired end	Mapping rate (%)	Non-conversion rate (%)
BAXF003A	WT	Tapetum	BS-seq	50.03	100	Single	60.9	0.63
BAXF003B	WT	Tapetum	BS-seq	44.74	100	Single	66.0	0.59
BAXF003C	WT	Tapetum	BS-seq	42.03	100	Single	59.8	0.67
BAXF010A	*rdr2*	Tapetum	BS-seq	72.02	75	Single	57.8	0.62
BAXF010B	*rdr2*	Tapetum	BS-seq	31.99	75	Single	63.9	0.69
BAXF010C	*rdr2*	Tapetum	BS-seq	12.89	75	Single	62.3	0.63
BAXF011A	*rdr6*	Tapetum	BS-seq	41.23	75	Single	19.2	0.66
BAXF011B	*rdr6*	Tapetum	BS-seq	51.57	75	Single	20.1	0.65
BAXF011C	*rdr6*	Tapetum	BS-seq	38.84	75	Single	23.7	0.69
HGXF015E	*pA9::RDR2 rdr2* #1-3	Meiocyte	BS-seq	50.82	75	Single	84.3	0.45
HGXF015F	*pA9::RDR2 rdr2* #2-4	Sperm	BS-seq	43.43	75	Single	88.1	0.88
JWXF002D	*pA9::RDR2 rdr2* #2-4	Meiocyte	BS-seq	54.49	100	Single	83.1	0.33
BAXF003D	WT	Tapetum	mRNA	97.28	50	Single	88.5	N/A
BAXF003E	WT	Tapetum	mRNA	84.11	50	Single	92.0	N/A
BAXF003F	WT	Tapetum	mRNA	85.24	50	Single	91.7	N/A
BAXF007C	WT	Tapetum	Total RNA	28.88	75	Single	73.5	N/A
HGXF013B	WT	Tapetum	Total RNA	18.14	75	Single	67.5	N/A
HGXF013C	WT	Tapetum	Total RNA	17.18	75	Single	69.3	N/A
HGXF014F	*rdr2*	Tapetum	mRNA	8.45	38	Paired	76.5	N/A
HGXF014G	*rdr2*	Tapetum	mRNA	9.14	38	Paired	68.2	N/A
HGXF014H	*rdr2*	Tapetum	mRNA	8.64	38	Paired	73.6	N/A
BAXF008	WT	Tapetum	Single cell (272)	300.00	100	Paired	Variable	N/A

Table A1: Table of sequencing libraries produced in this thesis. For detailed information about cell isolation, genotypes and library production see chapter methods. For BS-seq libraries non-conversion rate was calculated as the methylation level (C/C+T)) across the whole chloroplast genome, which is unmethylated in vivo. BAXF008 represents the pool of single-cell mRNA libraries sequenced.

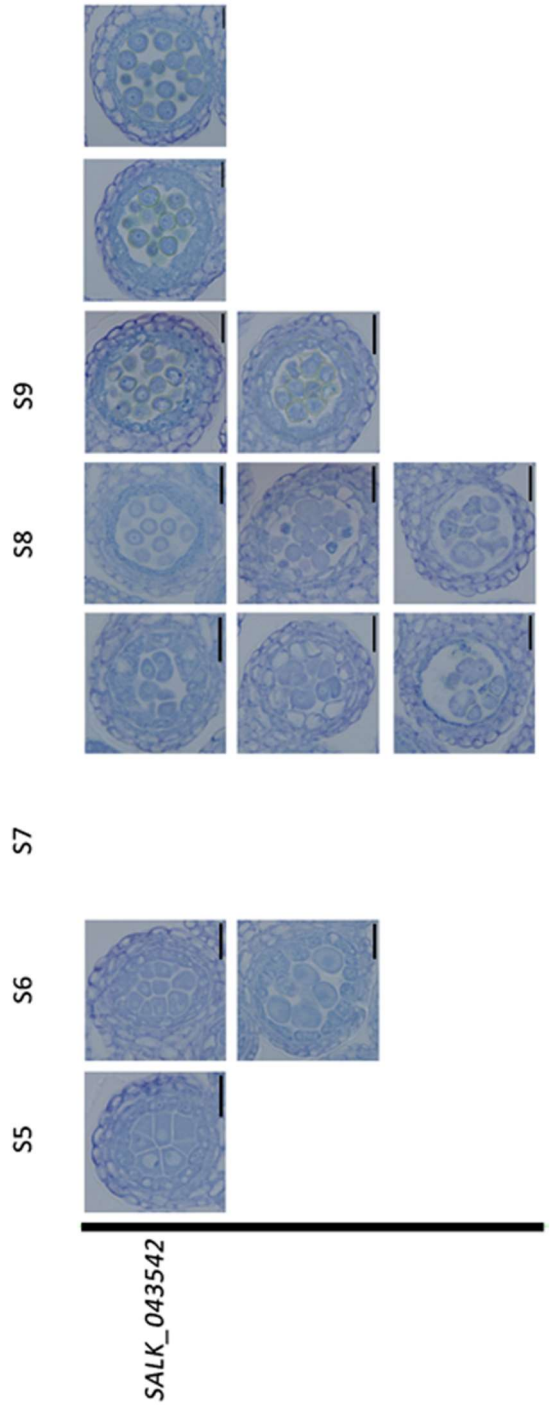


Figure A1: Transverse sectioning phenotype of a gt-3a mutant. gt-3a plants show tapetum defects earlier than other mutants examined. From stage 6 expanded meiocytes can be seen in some locules. At stage 7 a range of phenotypes are seen including: aborted microspores, vacuolated, degenerated, and multi-layered tapetum. General misshapen pollen grains are seen at stage 9. The phenotype was highly variable with some locules also appearing WT.

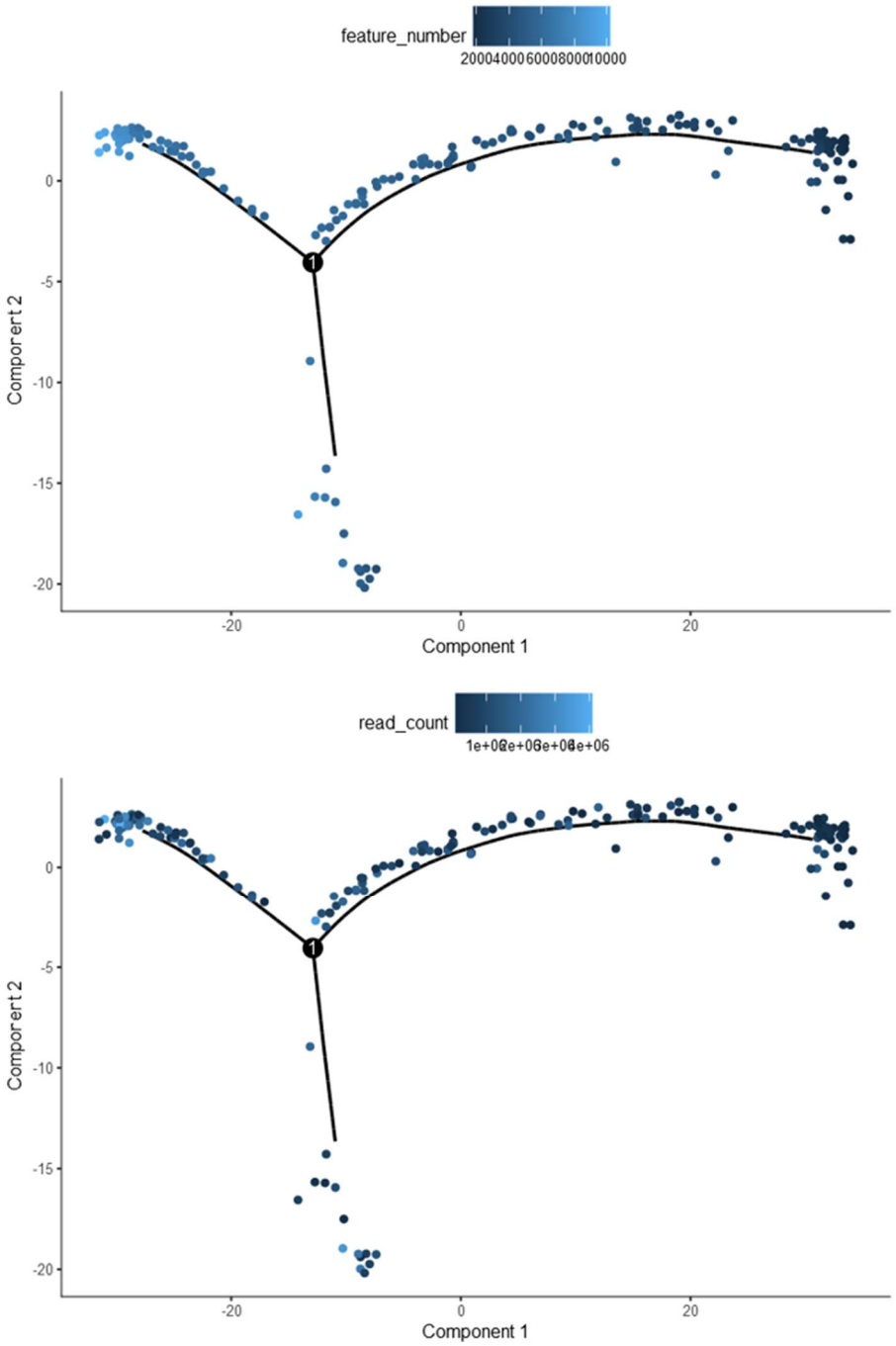


Figure A2: MST produced by monocle coloured by the number of genes expressed (feature_number) or read count. Monocle has separated cells based on sequencing depth, leading to a spurious inference of pseudotime.

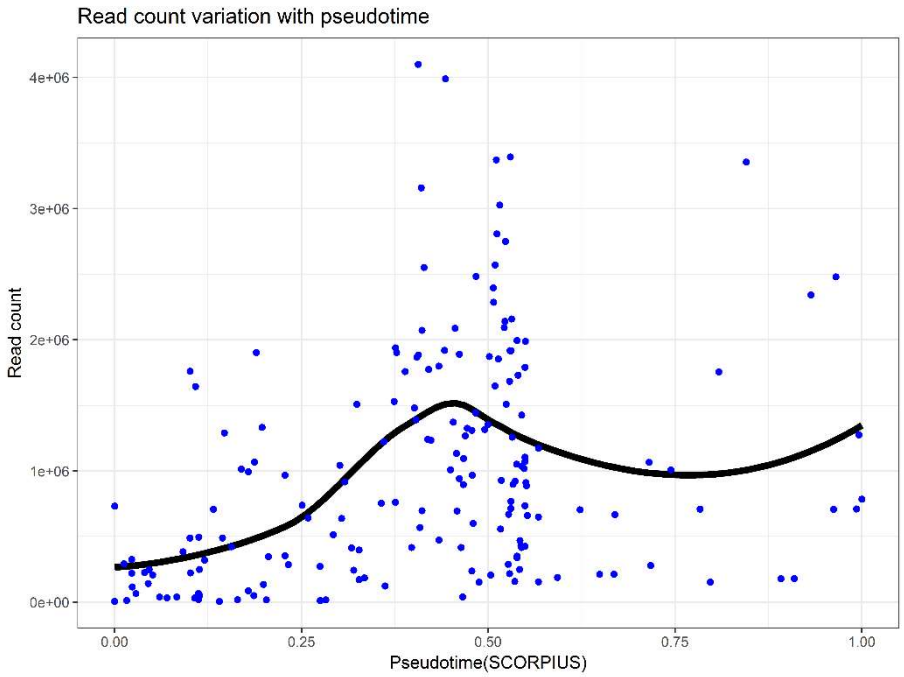


Figure A3: Read count variation with pseudotime inferred by SCORPIUS. Blue dots represent the read count in individual cells and the black line is a LOESS fitted curve.

Expression of differentially expressed genes within 500bp of Tapetum-specific hypermethylated DMRs							
Gene ID	Gene Name	WT Tapetum FPKM	*rdr2* Tapetum FPKM	log2 fold change	*p*-value	Distance to Nearest DMR	Number within 500bp
AT2G07748	AT2G07748	0.00	2020.15	10.98	0.00005	0	1
AT3G11930	AT3G11930	0.00	48.92	5.64	0.00005	0	1
AT3G22142	AT3G22142	0.00	35.70	5.20	0.00005	0	2
AT2G15320	AT2G15320	0.00	14.21	3.93	0.00005	0	1
AT1G55360	AT1G54920	0.00	13.43	3.85	0.00005	378	1
AT5G10310	AT5G10310	0.00	10.81	3.56	0.00005	0	1
AT2G07678	AT2G07678	2.39	23.82	2.87	0.0028	99	1
AT4G39940	APK2	0.00	6.15	2.84	0.00005	155	1
AT4G04830	MSRB5	0.00	6.06	2.82	0.0035	435	1
AT3G17040	HCF107	0.00	4.46	2.45	0.00005	0	1
AT5G27770	AT5G27770	154.14	838.13	2.44	0.00005	450	1
AT5G66590	AT5G66590	0.00	4.38	2.43	0.00005	0	1
AT1G64380	AT1G64380	0.00	2.97	1.99	0.00005	0	1
AT5G27330	AT5G27330	2.27	11.76	1.97	0.00135	0	1
AT4G13830	J20	17.83	71.41	1.94	0.00225	0	1
AT3G54320	WRI1	0.00	2.81	1.93	0.00025	0	1
AT2G36190	cwINV4	0.00	2.66	1.87	0.00005	0	1
AT3G23450	AT3G23450	0.00	2.65	1.87	0.00005	0	1
AT5G44710	AT5G44710	37.55	129.43	1.76	0.0015	468	1
AT2G07340	PFD1	64.06	217.23	1.75	0.0002	498	1
AT2G36780	AT2G36780	0.00	2.26	1.70	0.00005	0	1
AT5G27700	AT5G27700	389.60	1259.03	1.69	0.00005	246	1
AT4G28950	ROP9	0.00	2.05	1.61	0.00005	142	1
AT2G30070	KUP1	29.07	85.08	1.52	0.00005	0	1
AT4G17880	AT4G17880	0.00	1.79	1.48	0.00005	0	1
AT1G26630	ELF5A-2	964.48	2504.03	1.38	0.0032	382	1
AT5G46030	AT5G46030	117.14	295.91	1.33	0.0003	0	1
AT1G31163	AT1G31163	0.00	1.41	1.27	0.00005	198	1
AT2G05914	AT2G05642	0.00	1.00	1.00	0.00005	0	2
AT5G46210	CUL4	74.96	30.03	-1.29	0.00005	0	1
AT5G58130	ROS3	29.72	10.24	-1.45	0.00365	350	1
AT5G19770	TUA3	95.19	30.71	-1.60	0.00005	0	1
AT1G43670	AT1G43670	219.51	55.58	-1.96	0.00005	402	1
AT5G37770	CML24	708.90	172.14	-2.04	0.00005	0	1
AT5G28530	FRS10	10.07	0.90	-2.55	0.0026	0	1

Table A2: Table of genes within 500bp of tapetum-specific hypermethylated DMRs that are significantly differentially expressed between WT and rdr2 tapetum.

Bibliography

1. Leyser, O., and Day, S. (2009). Mechanisms in Plant Development, (Wiley).

2. Coen, E.S., and Meyerowitz, E.M. (1991). The war of the whorls: genetic interactions controlling flower development. Nature *353*, 31-37.

3. O'Maoileidigh, D.S., Graciet, E., and Wellmer, F. (2014). Gene networks controlling Arabidopsis thaliana flower development. New Phytol *201*, 16-30.

4. Ditta, G., Pinyopich, A., Robles, P., Pelaz, S., and Yanofsky, M.F. (2004). The SEP4 gene of Arabidopsis thaliana functions in floral organ and meristem identity. Curr Biol *14*, 1935-1940.

5. Pelaz, S., Ditta, G.S., Baumann, E., Wisman, E., and Yanofsky, M.F. (2000). B and C floral organ identity functions require SEPALLATA MADS-box genes. Nature *405*, 200-203.

6. Walbot, V. (1985). On the Life Strategies of Plants and Animals. Trends Genet *1*, 165-169.

7. Lanfear, R. (2018). Do plants have a segregated germline? PLoS Biol *16*, e2005439.

8. Nakajima, K. (2018). Be my baby: patterning toward plant germ cells. Curr Opin Plant Biol *41*, 110-115.

9. Walker, J., Gao, H., Zhang, J., Aldridge, B., Vickers, M., Higgins, J.D., and Feng, X. (2018). Sexual-lineage-specific DNA methylation regulates meiosis in Arabidopsis. Nat Genet *50*, 130-137.

10. Sanders, P.M., Bui, A.Q., Weterings, K., McIntire, K.N., Hsu, Y.C., Lee, P.Y., Truong, M.T., Beals, T.P., and Goldberg, R.B. (1999). Anther developmental defects in Arabidopsis thaliana male-sterile mutants. Sex Plant Reprod *11*, 297-322.

11. Quilichini, T.D., Douglas, C.J., and Samuels, A.L. (2014). New views of tapetum ultrastructure and pollen exine development in Arabidopsis thaliana. Ann Bot *114*, 1189-1201.

12. Zhang, D., Luo, X., and Zhu, L. (2011). Cytological analysis and genetic control of rice anther development. J Genet Genomics *38*, 379-390.

13. Poethig, R.S. (1987). Clonal Analysis of Cell Lineage Patterns in Plant Development. Am J Bot *74*, 581-594.

14. Zhang, D., and Yang, L. (2014). Specification of tapetum and microsporocyte cells within the anther. Curr Opin Plant Biol *17*, 49-55.

15. Steer, M.W. (1977). Differentiation of the tapetum in Avena. I. The cell surface. J Cell Sci *25*, 125-138.

16. Weiss, H., and Maluszynska, J. (2001). Molecular cytogenetic analysis of polyploidization in the anther tapetum of diploid and autotetraploid Arabidopsis thaliana plants. Ann Bot *87*, 729-735.

17. Hird, D.L., Worrall, D., Hodge, R., Smartt, S., Paul, W., and Scott, R. (1993). The anther-specific protein encoded by the Brassica napus and Arabidopsis thaliana A6 gene displays similarity to beta-1,3-glucanases. Plant J. *4*, 1023-1033.

18. Parish, R.W., and Li, S.F. (2010). Death of a tapetum: A programme of developmental altruism. Plant Sci *178*, 73-89.

19. Feng, X., and Dickinson, H.G. (2007). Packaging the male germline in plants. Trends Genet *23*, 503-510.

20. Kelliher, T., and Walbot, V. (2012). Hypoxia triggers meiotic fate acquisition in maize. Science *337*, 345-348.

21. Xing, S., and Zachgo, S. (2008). ROXY1 and ROXY2, two Arabidopsis glutaredoxin genes, are required for anther development. Plant J. *53*, 790-801.

22. Kelliher, T., and Walbot, V. (2014). Maize germinal cell initials accommodate hypoxia and precociously express meiotic genes. Plant J. *77*, 639-652.

23. Hong, L., Tang, D., Zhu, K., Wang, K., Li, M., and Cheng, Z. (2012). Somatic and reproductive cell development in rice anther is regulated by a putative glutaredoxin. Plant Cell *24*, 577-588.

24. Kelliher, T., Egger, R.L., Zhang, H., and Walbot, V. (2014). Unresolved issues in pre-meiotic anther development. Front Plant Sci *5*, 347.

25. Yang, W.C., Ye, D., Xu, J., and Sundaresan, V. (1999). The SPOROCYTELESS gene of Arabidopsis is required for initiation of sporogenesis and encodes a novel nuclear protein. Genes Dev *13*, 2108-2117.

26. Hord, C.L., Chen, C., Deyoung, B.J., Clark, S.E., and Ma, H. (2006). The BAM1/BAM2 receptor-like kinases are important regulators of *Arabidopsis* early anther development. Plant Cell *18*, 1667-1680.

27. Canales, C., Bhatt, A.M., Scott, R., and Dickinson, H. (2002). EXS, a putative LRR receptor kinase, regulates male germline cell number and tapetal identity and promotes seed development in Arabidopsis. Curr Biol *12*, 1718-1727.

28. Zhao, D.Z., Wang, G.F., Speal, B., and Ma, H. (2002). The excess microsporocytes1 gene encodes a putative leucine-rich repeat receptor protein kinase that controls somatic and reproductive cell fates in the Arabidopsis anther. Genes Dev *16*, 2021-2031.

29. Colcombet, J., Boisson-Dernier, A., Ros-Palau, R., Vera, C.E., and Schroeder, J.I. (2005). *Arabidopsis* SOMATIC EMBRYOGENESIS RECEPTOR KINASES1 and 2 are essential for tapetum development and microspore maturation. Plant Cell *17*, 3350-3361.

30. Albrecht, C., Russinova, E., Hecht, V., Baaijens, E., and de Vries, S. (2005). The Arabidopsis thaliana SOMATIC EMBRYOGENESIS RECEPTOR-LIKE KINASES1 and 2 control male sporogenesis. Plant Cell *17*, 3337-3349.

31. Yang, S.L., Xie, L.F., Mao, H.Z., Puah, C.S., Yang, W.C., Jiang, L., Sundaresan, V., and Ye, D. (2003). Tapetum determinant1 is required for cell specialization in the Arabidopsis anther. Plant Cell *15*, 2792-2804.

32. Yang, S.L., Jiang, L., Puah, C.S., Xie, L.F., Zhang, X.Q., Chen, L.Q., Yang, W.C., and Ye, D. (2005). Overexpression of TAPETUM DETERMINANT1 alters the cell fates in the Arabidopsis carpel and tapetum via genetic interaction with excess microsporocytes1/extra sporogenous cells. Plant Physiol *139*, 186-191.

33. Jia, G., Liu, X., Owen, H.A., and Zhao, D. (2008). Signaling of cell fate determination by the TPD1 small protein and EMS1 receptor kinase. Proc Natl Acad Sci U S A *105*, 2220-2225.

34. Feng, X., and Dickinson, H.G. (2010). Tapetal cell fate, lineage and proliferation in the Arabidopsis anther. Development *137*, 2409-2416.

35. De Veylder, L., Larkin, J.C., and Schnittger, A. (2011). Molecular control and function of endoreplication in development and physiology. Trends Plant Sci *16*, 624-634.

36. Witkus, E.R. (1945). Endomitotic Tapetal Cell Divisions in Spinacia. Am J Bot *32*, 326-330.

37. Carvalheira, G., and Guerra, M. (1994). The Polytene Chromosomes of Anther Tapetum of Some Phaseolus Species. Cytologia *59*, 211-217.

38. Lima-De-Faria, A., Pero, R., Avanzi, S., Durante, M., Ståhle, U., D'Amato, F., and Granström, H. (2009). Relation between ribosomal RNA genes and the DNA satellites of Phaseolus coccineus. Hereditas *79*, 5-19.

39. Shu, Z., Row, S., and Deng, W.M. (2018). Endoreplication: The Good, the Bad, and the Ugly. Trends Cell Biol *28*, 465-474.

40. Heslop-Harrison, J. (1968). Pollen wall development. The succession of events in the growth of intricately patterned pollen walls is described and discussed. Science *161*, 230-237.

41. Quilichini, T.D., Grienenberger, E., and Douglas, C.J. (2015). The biosynthesis, composition and assembly of the outer pollen wall: A tough case to crack. Phytochemistry *113*, 170-182.

42. Jiang, J., Zhang, Z., and Cao, J. (2013). Pollen wall development: the associated enzymes and metabolic pathways. Plant Biol *15*, 249-263.

43. Ivanov, R., Fobis-Loisy, I., and Gaude, T. (2010). When no means no: guide to Brassicaceae self-incompatibility. Trends Plant Sci *15*, 387-394.

44. Rhee, S.Y., and Somerville, C.R. (1998). Tetrad pollen formation in quartet mutants of Arabidopsis thaliana is associated with persistence of pectic polysaccharides of the pollen mother cell wall. Plant J. *15*, 79-88.

45. Lu, P., Chai, M., Yang, J., Ning, G., Wang, G., and Ma, H. (2014). The Arabidopsis CALLOSE DEFECTIVE MICROSPORE1 gene is required for male fertility through regulating callose metabolism during microsporogenesis. Plant Physiol *164*, 1893-1904.

46. Xu, T., Zhang, C., Zhou, Q., and Yang, Z.N. (2016). Pollen wall pattern in Arabidopsis. Sci Bull *61*, 832-837.

47. Piffanelli, P., Ross, J.H.E., and Murphy, D.J. (1998). Biogenesis and function of the lipidic structures of pollen grains. Sex Plant Reprod *11*, 65-80.

48. de Azevedo Souza, C., Kim, S.S., Koch, S., Kienow, L., Schneider, K., McKim, S.M., Haughn, G.W., Kombrink, E., and Douglas, C.J. (2009). A novel fatty Acyl-CoA Synthetase is required for pollen development and sporopollenin biosynthesis in Arabidopsis. Plant Cell *21*, 507-525.

49. Dobritsa, A.A., Lei, Z., Nishikawa, S., Urbanczyk-Wochniak, E., Huhman, D.V., Preuss, D., and Sumner, L.W. (2010). LAP5 and LAP6 encode anther-specific proteins with similarity to chalcone synthase essential for pollen exine development in Arabidopsis. Plant Physiol *153*, 937-955.

50. Grienenberger, E., Kim, S.S., Lallemand, B., Geoffroy, P., Heintz, D., Souza Cde, A., Heitz, T., Douglas, C.J., and Legrand, M. (2010). Analysis of TETRAKETIDE alpha-PYRONE REDUCTASE function in Arabidopsis thaliana reveals a previously unknown, but conserved, biochemical pathway in sporopollenin monomer biosynthesis. Plant Cell *22*, 4067-4083.

51. Morant, M., Jorgensen, K., Schaller, H., Pinot, F., Moller, B.L., Werck-Reichhart, D., and Bak, S. (2007). CYP703 is an ancient cytochrome P450 in

land plants catalyzing in-chain hydroxylation of lauric acid to provide building blocks for sporopollenin synthesis in pollen. Plant Cell *19*, 1473-1487.

52. Dobritsa, A.A., Shrestha, J., Morant, M., Pinot, F., Matsuno, M., Swanson, R., Moller, B.L., and Preuss, D. (2009). CYP704B1 is a long-chain fatty acid omega-hydroxylase essential for sporopollenin synthesis in pollen of Arabidopsis. Plant Physiol *151*, 574-589.

53. Aarts, M.G., Hodge, R., Kalantidis, K., Florack, D., Wilson, Z.A., Mulligan, B.J., Stiekema, W.J., Scott, R., and Pereira, A. (1997). The Arabidopsis MALE STERILITY 2 protein shares similarity with reductases in elongation/condensation complexes. Plant J. *12*, 615-623.

54. Rowland, O., Lee, R., Franke, R., Schreiber, L., and Kunst, L. (2007). The CER3 wax biosynthetic gene from Arabidopsis thaliana is allelic to WAX2/YRE/FLP1. FEBS Lett *581*, 3538-3544.

55. Bourdenx, B., Bernard, A., Domergue, F., Pascal, S., Leger, A., Roby, D., Pervent, M., Vile, D., Haslam, R.P., Napier, J.A., et al. (2011). Overexpression of Arabidopsis ECERIFERUM1 promotes wax very-long-chain alkane biosynthesis and influences plant response to biotic and abiotic stresses. Plant Physiol *156*, 29-45.

56. Aarts, M.G., Keijzer, C.J., Stiekema, W.J., and Pereira, A. (1995). Molecular characterization of the CER1 gene of arabidopsis involved in epicuticular wax biosynthesis and pollen fertility. Plant Cell *7*, 2115-2127.

57. Fellenberg, C., and Vogt, T. (2015). Evolutionarily conserved phenylpropanoid pattern on angiosperm pollen. Trends Plant Sci *20*, 212-218.

58. Zhao, Q. (2016). Lignification: Flexibility, Biosynthesis and Regulation. Trends Plant Sci *21*, 713-721.

59. Schilmiller, A.L., Stout, J., Weng, J.K., Humphreys, J., Ruegger, M.O., and Chapple, C. (2009). Mutations in the cinnamate 4-hydroxylase gene impact metabolism, growth and development in Arabidopsis. Plant J. *60*, 771-782.

60. Dobritsa, A.A., Nishikawa, S., Preuss, D., Urbanczyk-Wochniak, E., Sumner, L.W., Hammond, A., Carlson, A.L., and Swanson, R.J. (2009). LAP3, a novel plant protein required for pollen development, is essential for proper exine formation. Sex Plant Reprod *22*, 167-177.

61. Quilichini, T.D., Samuels, A.L., and Douglas, C.J. (2014). ABCG26-mediated polyketide trafficking and hydroxycinnamoyl spermidines contribute to pollen wall exine formation in Arabidopsis. Plant Cell *26*, 4483-4498.

62. Huang, M.D., Chen, T.L., and Huang, A.H. (2013). Abundant type III lipid transfer proteins in Arabidopsis tapetum are secreted to the locule and become a constituent of the pollen exine. Plant Physiol *163*, 1218-1229.

63. Huang, M.D., Wei, F.J., Wu, C.C., Hsing, Y.I., and Huang, A.H. (2009). Analyses of advanced rice anther transcriptomes reveal global tapetum secretory functions and potential proteins for lipid exine formation. Plant Physiol *149*, 694-707.

64. Rubinelli, P., Hu, Y., and Ma, H. (1998). Identification, sequence analysis and expression studies of novel anther-specific genes of Arabidopsis thaliana. Plant Mol Biol *37*, 607-619.

65. Paul, W., Hodge, R., Smartt, S., Draper, J., and Scott, R. (1992). The isolation and characterisation of the tapetum-specific Arabidopsis thaliana A9 gene. Plant Mol Biol *19*, 611-622.

66. Hsieh, K., and Huang, A.H. (2007). Tapetosomes in Brassica tapetum accumulate endoplasmic reticulum-derived flavonoids and alkanes for delivery to the pollen surface. Plant Cell *19*, 582-596.

67. Hsieh, K., and Huang, A.H. (2005). Lipid-rich tapetosomes in Brassica tapetum are composed of oleosin-coated oil droplets and vesicles, both assembled in and then detached from the endoplasmic reticulum. Plant J. *43*, 889-899.

68. Hsieh, K., and Huang, A.H. (2004). Endoplasmic reticulum, oleosins, and oils in seeds and tapetum cells. Plant Physiol *136*, 3427-3434.

69. Van Hautegem, T., Waters, A.J., Goodrich, J., and Nowack, M.K. (2015). Only in dying, life: programmed cell death during plant development. Trends Plant Sci *20*, 102-113.

70. Aya, K., Ueguchi-Tanaka, M., Kondo, M., Hamada, K., Yano, K., Nishimura, M., and Matsuoka, M. (2009). Gibberellin modulates anther development in rice via the transcriptional regulation of GAMYB. Plant Cell *21*, 1453-1472.

71. Millar, A.A., and Gubler, F. (2005). The Arabidopsis GAMYB-like genes, MYB33 and MYB65, are microRNA-regulated genes that redundantly facilitate anther development. Plant Cell *17*, 705-721.

72. Plackett, A.R., Ferguson, A.C., Powers, S.J., Wanchoo-Kohli, A., Phillips, A.L., Wilson, Z.A., Hedden, P., and Thomas, S.G. (2014). DELLA activity is required for successful pollen development in the Columbia ecotype of Arabidopsis. New Phytol *201*, 825-836.

73. Zhang, D., Liu, D., Lv, X., Wang, Y., Xun, Z., Liu, Z., Li, F., and Lu, H. (2014). The cysteine protease CEP1, a key executor involved in tapetal programmed cell death, regulates pollen development in Arabidopsis. Plant Cell *26*, 2939-2961.

74. Phan, H.A., Iacuone, S., Li, S.F., and Parish, R.W. (2011). The MYB80 transcription factor is required for pollen development and the regulation of tapetal programmed cell death in Arabidopsis thaliana. Plant Cell *23*, 2209-2224.

75. Balk, J., and Leaver, C.J. (2001). The PET1-CMS mitochondrial mutation in sunflower is associated with premature programmed cell death and cytochrome c release. Plant Cell *13*, 1803-1818.

76. Kurusu, T., and Kuchitsu, K. (2017). Autophagy, programmed cell death and reactive oxygen species in sexual reproduction in plants. J Plant Res *130*, 491-499.

77. Hu, L., Liang, W., Yin, C., Cui, X., Zong, J., Wang, X., Hu, J., and Zhang, D. (2011). Rice MADS3 regulates ROS homeostasis during late anther development. Plant Cell *23*, 515-533.

78. Xie, H.T., Wan, Z.Y., Li, S., and Zhang, Y. (2014). Spatiotemporal Production of Reactive Oxygen Species by NADPH Oxidase Is Critical for Tapetal Programmed Cell Death and Pollen Development in Arabidopsis. Plant Cell *26*, 2007-2023.

79. Zhang, W., Sun, Y., Timofejeva, L., Chen, C., Grossniklaus, U., and Ma, H. (2006). Regulation of *Arabidopsis* tapetum development and function by DYSFUNCTIONAL TAPETUM1 (DYT1) encoding a putative bHLH transcription factor. Development *133*, 3085-3095.

80. Zhu, J., Lou, Y., Xu, X., and Yang, Z.N. (2011). A genetic pathway for tapetum development and function in Arabidopsis. J Integr Plant Biol *53*, 892-900.

81. Feng, B., Lu, D., Ma, X., Peng, Y., Sun, Y., Ning, G., and Ma, H. (2012). Regulation of the Arabidopsis anther transcriptome by DYT1 for pollen development. Plant J. *72*, 612-624.

82. Zhu, E., You, C., Wang, S., Cui, J., Niu, B., Wang, Y., Qi, J., Ma, H., and Chang, F. (2015). The DYT1-interacting proteins bHLH010, bHLH089 and bHLH091 are redundantly required for Arabidopsis anther development and transcriptome. Plant J. *83*, 976-990.

83. Gu, J.N., Zhu, J., Yu, Y., Teng, X.D., Lou, Y., Xu, X.F., Liu, J.L., and Yang, Z.N. (2014). DYT1 directly regulates the expression of TDF1 for tapetum development and pollen wall formation in Arabidopsis. Plant J. *80*, 1005-1013.

84. Zhu, J., Chen, H., Li, H., Gao, J.F., Jiang, H., Wang, C., Guan, Y.F., and Yang, Z.N. (2008). Defective in Tapetal development and function 1 is essential for anther development and tapetal function for microspore maturation in Arabidopsis. Plant J. *55*, 266-277.

85. Li, D.D., Xue, J.S., Zhu, J., and Yang, Z.N. (2017). Gene Regulatory Network for Tapetum Development in Arabidopsis thaliana. Front Plant Sci *8*, 1559.

86. Lou, Y., Zhou, H.S., Han, Y., Zeng, Q.Y., Zhu, J., and Yang, Z.N. (2018). Positive regulation of AMS by TDF1 and the formation of a TDF1-AMS complex are required for anther development in Arabidopsis thaliana. New Phytol *217*, 378-391.

87. Ferguson, A.C., Pearce, S., Band, L.R., Yang, C., Ferjentsikova, I., King, J., Yuan, Z., Zhang, D., and Wilson, Z.A. (2017). Biphasic regulation of the transcription factor ABORTED MICROSPORES (AMS) is essential for tapetum and pollen development in Arabidopsis. New Phytol *213*, 778-790.

88. Li, S.F., Higginson, T., and Parish, R.W. (1999). A novel MYB-related gene from Arabidopsis thaliana expressed in developing anthers. Plant Cell Physiol *40*, 343-347.

89. Wang, K., Guo, Z.L., Zhou, W.T., Zhang, C., Zhang, Z.Y., Lou, Y., Xiong, S.X., Yao, X.Z., Fan, J.J., Zhu, J., et al. (2018). The Regulation of Sporopollenin Biosynthesis Genes for Rapid Pollen Wall Formation. Plant Physiol *178*, 283-294.

90. Zhang, Z.B., Zhu, J., Gao, J.F., Wang, C., Li, H., Li, H., Zhang, H.Q., Zhang, S., Wang, D.M., Wang, Q.X., et al. (2007). Transcription factor AtMYB103 is required for anther development by regulating tapetum development, callose dissolution and exine formation in Arabidopsis. Plant J. *52*, 528-538.

91. Xiong, S.X., Lu, J.Y., Lou, Y., Teng, X.D., Gu, J.N., Zhang, C., Shi, Q.S., Yang, Z.N., and Zhu, J. (2016). The transcription factors MS188 and AMS form a complex to activate the expression of CYP703A2 for sporopollenin biosynthesis in Arabidopsis thaliana. Plant J. *88*, 936-946.

92. Alves-Ferreira, M., Wellmer, F., Banhara, A., Kumar, V., Riechmann, J.L., and Meyerowitz, E.M. (2007). Global expression profiling applied to the analysis of Arabidopsis stamen development. Plant Physiol *145*, 747-762.

93. Preston, J., Wheeler, J., Heazlewood, J., Li, S.F., and Parish, R.W. (2004). AtMYB32 is required for normal pollen development in Arabidopsis thaliana. Plant J. *40*, 979-995.

94. Wilson, Z.A., Morroll, S.M., Dawson, J., Swarup, R., and Tighe, P.J. (2001). The Arabidopsis MALE STERILITY1 (MS1) gene is a transcriptional regulator of

male gametogenesis, with homology to the PHD-finger family of transcription factors. Plant J. *28*, 27-39.

95. Vizcay-Barrena, G., and Wilson, Z.A. (2006). Altered tapetal PCD and pollen wall development in the Arabidopsis ms1 mutant. J Exp Bot *57*, 2709-2717.

96. Yang, C., Vizcay-Barrena, G., Conner, K., and Wilson, Z.A. (2007). MALE STERILITY1 is required for tapetal development and pollen wall biosynthesis. Plant Cell *19*, 3530-3548.

97. Ito, T., and Shinozaki, K. (2002). The MALE STERILITY1 gene of Arabidopsis, encoding a nuclear protein with a PHD-finger motif, is expressed in tapetal cells and is required for pollen maturation. Plant Cell Physiol *43*, 1285-1292.

98. McClintock, B. (1950). The origin and behavior of mutable loci in maize. Proc Natl Acad Sci U S A *36*, 344-355.

99. Underwood, C.J., Henderson, I.R., and Martienssen, R.A. (2017). Genetic and epigenetic variation of transposable elements in Arabidopsis. Curr Opin Plant Biol *36*, 135-141.

100. Joly-Lopez, Z., and Bureau, T.E. (2014). Diversity and evolution of transposable elements in Arabidopsis. Chromosome Res *22*, 203-216.

101. Singer, T., Yordan, C., and Martienssen, R.A. (2001). Robertson's Mutator transposons in A. thaliana are regulated by the chromatin-remodeling gene Decrease in DNA Methylation (DDM1). Genes Dev *15*, 591-602.

102. Kapitonov, V.V., and Jurka, J. (2001). Rolling-circle transposons in eukaryotes. Proc Natl Acad Sci U S A *98*, 8714-8719.

103. Liu, S., Yeh, C.T., Ji, T., Ying, K., Wu, H., Tang, H.M., Fu, Y., Nettleton, D., and Schnable, P.S. (2009). Mu transposon insertion sites and meiotic recombination events co-localize with epigenetic marks for open chromatin across the maize genome. PLoS Genet *5*, e1000733.

104. Cavrak, V.V., Lettner, N., Jamge, S., Kosarewicz, A., Bayer, L.M., and Mittelsten Scheid, O. (2014). How a Retrotransposon Exploits the Plant's Heat Stress Response for Its Activation. PLoS Genet *10*, e1004115.

105. Martinez, G., and Slotkin, R.K. (2012). Developmental relaxation of transposable element silencing in plants: functional or byproduct? Curr Opin Plant Biol *15*, 496-502.

106. Chen, C., Farmer, A.D., Langley, R.J., Mudge, J., Crow, J.A., May, G.D., Huntley, J., Smith, A.G., and Retzel, E.F. (2010). Meiosis-specific gene discovery in plants: RNA-Seq applied to isolated Arabidopsis male meiocytes. BMC Plant Biol *10*, 280.

107. Yang, H., Lu, P., Wang, Y., and Ma, H. (2011). The transcriptome landscape of Arabidopsis male meiocytes from high-throughput sequencing: the complexity and evolution of the meiotic process. Plant J. *65*, 503-516.

108. 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.

109. Martinez, G., Panda, K., Kohler, C., and Slotkin, R.K. (2016). Silencing in sperm cells is directed by RNA movement from the surrounding nurse cell. Nat Plants *2*, 16030.

110. Luger, K., Mader, A.W., Richmond, R.K., Sargent, D.F., and Richmond, T.J. (1997). Crystal structure of the nucleosome core particle at 2.8 A resolution. Nature *389*, 251-260.

111. Roudier, F., Ahmed, I., Berard, C., Sarazin, A., Mary-Huard, T., Cortijo, S., Bouyer, D., Caillieux, E., Duvernois-Berthet, E., Al-Shikhley, L., et al. (2011). Integrative epigenomic mapping defines four main chromatin states in Arabidopsis. EMBO J *30*, 1928-1938.

112. Jackson, J.P., Lindroth, A.M., Cao, X., and Jacobsen, S.E. (2002). Control of CpNpG DNA methylation by the KRYPTONITE histone H3 methyltransferase. Nature *416*, 556-560.

113. Ebbs, M.L., and Bender, J. (2006). Locus-specific control of DNA methylation by the *Arabidopsis* SUVH5 histone methyltransferase. Plant Cell *18*, 1166-1176.

114. Bernatavichute, Y.V., Zhang, X., Cokus, S., Pellegrini, M., and Jacobsen, S.E. (2008). Genome-wide association of histone H3 lysine nine methylation with CHG DNA methylation in Arabidopsis thaliana. PLoS One *3*, e3156.

115. Margueron, R., and Reinberg, D. (2010). Chromatin structure and the inheritance of epigenetic information. Nat Rev Genet *11*, 285-296.

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

117. Happel, N., and Doenecke, D. (2009). Histone H1 and its isoforms: contribution to chromatin structure and function. Gene *431*, 1-12.

118. Zemach, A., Kim, M.Y., Hsieh, P.H., Coleman-Derr, D., Eshed-Williams, L., Thao, K., Harmer, S.L., and Zilberman, D. (2013). The Arabidopsis nucleosome remodeler DDM1 allows DNA methyltransferases to access H1-containing heterochromatin. Cell *153*, 193-205.

119. Zhou, V.W., Goren, A., and Bernstein, B.E. (2011). Charting histone modifications and the functional organization of mammalian genomes. Nat Rev Genet *12*, 7-18.

120. Zhang, X., Yazaki, J., Sundaresan, A., Cokus, S., Chan, S.W., Chen, H., Henderson, I.R., Shinn, P., Pellegrini, M., Jacobsen, S.E., et al. (2006). Genome-wide high-resolution mapping and functional analysis of DNA methylation in arabidopsis. Cell *126*, 1189-1201.

121. Law, J.A., and Jacobsen, S.E. (2010). Establishing, maintaining and modifying DNA methylation patterns in plants and animals. Nat Rev Genet *11*, 204-220.

122. Vongs, A., Kakutani, T., Martienssen, R., and Richards, E. (1993). Arabidopsis thaliana DNA methylation mutants. Science *260*, 1926-1928.

123. Woo, H.R., Dittmer, T.A., and Richards, E.J. (2008). Three SRA-domain methylcytosine-binding proteins cooperate to maintain global CpG methylation and epigenetic silencing in Arabidopsis. PLoS Genet *4*, e1000156.

124. Du, J., Johnson, L.M., Jacobsen, S.E., and Patel, D.J. (2015). DNA methylation pathways and their crosstalk with histone methylation. Nat Rev Mol Cell Biol *16*, 519-532.

125. Du, J., Zhong, X., Bernatavichute, Y.V., Stroud, H., Feng, S., Caro, E., Vashisht, A.A., Terragni, J., Chin, H.G., Tu, A., et al. (2012). Dual binding of chromomethylase domains to H3K9me2-containing nucleosomes directs DNA methylation in plants. Cell *151*, 167-180.

126. Stroud, H., Do, T., Du, J., Zhong, X., Feng, S., Johnson, L., Patel, D.J., and Jacobsen, S.E. (2014). Non-CG methylation patterns shape the epigenetic landscape in Arabidopsis. Nat Struct Mol Biol *21*, 64-72.

127. Stroud, H., Greenberg, M.V., Feng, S., Bernatavichute, Y.V., and Jacobsen, S.E. (2013). Comprehensive analysis of silencing mutants reveals complex regulation of the Arabidopsis methylome. Cell *152*, 352-364.

128. Matzke, M.A., and Mosher, R.A. (2014). RNA-directed DNA methylation: an epigenetic pathway of increasing complexity. Nat Rev Genet *15*, 394-408.

129. Zhai, J., Bischof, S., Wang, H., Feng, S., Lee, T.F., Teng, C., Chen, X., Park, S.Y., Liu, L., Gallego-Bartolome, J., et al. (2015). A One Precursor One siRNA Model for Pol IV-Dependent siRNA Biogenesis. Cell *163*, 445-455.

130. Zhou, M., Palanca, A.M.S., and Law, J.A. (2018). Locus-specific control of the de novo DNA methylation pathway in Arabidopsis by the CLASSY family. Nat Genet *50*, 865-873.

131. Law, J.A., Du, J.M., Hale, C.J., Feng, S.H., Krajewski, K., Palanca, A.M.S., Strahl, B.D., Patel, D.J., and Jacobsen, S.E. (2013). Polymerase IV occupancy at RNA-directed DNA methylation sites requires SHH1. Nature *498*, 385-389.

132. Law, J.A., Vashisht, A.A., Wohlschlegel, J.A., and Jacobsen, S.E. (2011). SHH1, a homeodomain protein required for DNA methylation, as well as RDR2, RDM4, and chromatin remodeling factors, associate with RNA polymerase IV. PLoS Genet *7*, e1002195.

133. Haag, J.R., Ream, T.S., Marasco, M., Nicora, C.D., Norbeck, A.D., Pasa-Tolic, L., and Pikaard, C.S. (2012). *In vitro* transcription activities of Pol IV, Pol V, and RDR2 reveal coupling of Pol IV and RDR2 for dsRNA synthesis in plant RNA silencing. Mol Cell *48*, 811-818.

134. Ji, L., and Chen, X. (2012). Regulation of small RNA stability: methylation and beyond. Cell Res *22*, 624-636.

135. Greenberg, M.V.C., Ausin, I., Chan, S.W.L., Cokus, S.J., Cuperus, J.T., Feng, S., Law, J.A., Chu, C., Pellegrini, M., Carrington, J.C., et al. (2011). Identification of genes required for de novo DNA methylation in Arabidopsis. Epigenetics *6*, 344-354.

136. Chan, S.W., Zilberman, D., Xie, Z., Johansen, L.K., Carrington, J.C., and Jacobsen, S.E. (2004). RNA silencing genes control de novo DNA methylation. Science *303*, 1336.

137. Mallory, A., and Vaucheret, H. (2010). Form, function, and regulation of ARGONAUTE proteins. Plant Cell *22*, 3879-3889.

138. Eun, C., Lorkovic, Z.J., Naumann, U., Long, Q., Havecker, E.R., Simon, S.A., Meyers, B.C., Matzke, A.J., and Matzke, M. (2011). AGO6 functions in RNA-mediated transcriptional gene silencing in shoot and root meristems in Arabidopsis thaliana. PLoS One *6*, e25730.

139. Olmedo-Monfil, V., Duran-Figueroa, N., Arteaga-Vazquez, M., Demesa-Arevalo, E., Autran, D., Grimanelli, D., Slotkin, R.K., Martienssen, R.A., and Vielle-Calzada, J.P. (2010). Control of female gamete formation by a small RNA pathway in Arabidopsis. Nature *464*, 628-632.

140. Li, C.F., Pontes, O., El-Shami, M., Henderson, I.R., Bernatavichute, Y.V., Chan, S.W., Lagrange, T., Pikaard, C.S., and Jacobsen, S.E. (2006). An ARGONAUTE4-containing nuclear processing center colocalized with Cajal bodies in *Arabidopsis thaliana*. Cell *126*, 93-106.

141. Jullien, P.E., Susaki, D., Yelagandula, R., Higashiyama, T., and Berger, F. (2012). DNA methylation dynamics during sexual reproduction in Arabidopsis thaliana. Curr Biol *22*, 1825-1830.

142. Wada, Y., Ohya, H., Yamaguchi, Y., Koizumi, N., and Sano, H. (2003). Preferential de novo methylation of cytosine residues in non-CpG sequences by a domains rearranged DNA methyltransferase from tobacco plants. J Biol Chem *278*, 42386-42393.

143. Johnson, L.M., Du, J., Hale, C.J., Bischof, S., Feng, S., Chodavarapu, R.K., Zhong, X., Marson, G., Pellegrini, M., Segal, D.J., et al. (2014). SRA- and SET-domain-containing proteins link RNA polymerase V occupancy to DNA methylation. Nature *507*, 124-128.

144. Liu, Z.W., Shao, C.R., Zhang, C.J., Zhou, J.X., Zhang, S.W., Li, L., Chen, S., Huang, H.W., Cai, T., and He, X.J. (2014). The SET domain proteins SUVH2 and SUVH9 are required for Pol V occupancy at RNA-directed DNA methylation loci. PLoS Genet *10*, e1003948.

145. Nuthikattu, S., McCue, A.D., Panda, K., Fultz, D., DeFraia, C., Thomas, E.N., and Slotkin, R.K. (2013). The initiation of epigenetic silencing of active transposable elements is triggered by RDR6 and 21-22 nucleotide small interfering RNAs. Plant Physiol *162*, 116-131.

146. Wu, L., Mao, L., and Qi, Y. (2012). Roles of dicer-like and argonaute proteins in TAS-derived small interfering RNA-triggered DNA methylation. Plant Physiol *160*, 990-999.

147. Bond, D.M., and Baulcombe, D.C. (2015). Epigenetic transitions leading to heritable, RNA-mediated de novo silencing in Arabidopsis thaliana. Proc Natl Acad Sci U S A *112*, 917-922.

148. McCue, A.D., Panda, K., Nuthikattu, S., Choudury, S.G., Thomas, E.N., and Slotkin, R.K. (2015). ARGONAUTE 6 bridges transposable element mRNA-derived siRNAs to the establishment of DNA methylation. EMBO J *34*, 20-35.

149. Mari-Ordonez, A., Marchais, A., Etcheverry, M., Martin, A., Colot, V., and Voinnet, O. (2013). Reconstructing de novo silencing of an active plant retrotransposon. Nat Genet *45*, 1029-1039.

150. 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.

151. Kawakatsu, T., Stuart, T., Valdes, M., Breakfield, N., Schmitz, R.J., Nery, J.R., Urich, M.A., Han, X., Lister, R., Benfey, P.N., et al. (2016). Unique cell-type-specific patterns of DNA methylation in the root meristem. Nat Plants *2*, 16058.

152. Ibarra, C.A., Feng, X., Schoft, V.K., Hsieh, T.F., Uzawa, R., Rodrigues, J.A., Zemach, A., Chumak, N., Machlicova, A., Nishimura, T., et al. (2012). Active DNA demethylation in plant companion cells reinforces transposon methylation in gametes. Science *337*, 1360-1364.

153. Lewsey, M.G., Hardcastle, T.J., Melnyk, C.W., Molnar, A., Valli, A., Urich, M.A., Nery, J.R., Baulcombe, D.C., and Ecker, J.R. (2016). Mobile small RNAs regulate genome-wide DNA methylation. Proc Natl Acad Sci U S A *113*, E801-810.

154. Calarco, J.P., Borges, F., Donoghue, M.T., Van Ex, F., Jullien, P.E., Lopes, T., Gardner, R., Berger, F., Feijo, J.A., Becker, J.D., et al. (2012). Reprogramming of DNA methylation in pollen guides epigenetic inheritance via small RNA. Cell *151*, 194-205.

155. Rodrigues, J.A., Ruan, R., Nishimura, T., Sharma, M.K., Sharma, R., Ronald, P.C., Fischer, R.L., and Zilberman, D. (2013). Imprinted expression of genes and small RNA is associated with localized hypomethylation of the maternal genome in rice endosperm. Proc Natl Acad Sci U S A *110*, 7934-7939.

156. Zhai, J., Zhang, H., Arikit, S., Huang, K., Nan, G.L., Walbot, V., and Meyers, B.C. (2015). Spatiotemporally dynamic, cell-type-dependent premeiotic and meiotic phasiRNAs in maize anthers. Proc Natl Acad Sci U S A *112*, 3146-3151.

157. Gehring, M., Huh, J.H., Hsieh, T.F., Penterman, J., Choi, Y., Harada, J.J., Goldberg, R.B., and Fischer, R.L. (2006). DEMETER DNA glycosylase establishes MEDEA polycomb gene self-imprinting by allele-specific demethylation. Cell *124*, 495-506.

158. Dickinson, H., and Scott, R. (2002). DEMETER, Goddess of the harvest, activates maternal MEDEA to produce the perfect seed. Mol Cell *10*, 5-7.

159. Choi, Y.H., Gehring, M., Johnson, L., Hannon, M., Harada, J.J., Goldberg, R.B., Jacobsen, S.E., and Fischer, R.L. (2002). DEMETER, a DNA glycosylase domain protein, is required for endosperm gene imprinting and seed viability in Arabidopsis. Cell *110*, 33-42.

160. Morales-Ruiz, T., Ortega-Galisteo, A.P., Ponferrada-Marin, M.I., Martinez-Macias, M.I., Ariza, R.R., and Roldan-Arjona, T. (2006). *DEMETER* and *REPRESSOR OF SILENCING 1* encode 5-methylcytosine DNA glycosylases. Proc Natl Acad Sci U S A *103*, 6853-6858.

161. She, W., and Baroux, C. (2014). Chromatin dynamics during plant sexual reproduction. Front Plant Sci *5*, 354.

162. Kawashima, T., and Berger, F. (2014). Epigenetic reprogramming in plant sexual reproduction. Nat Rev Genet *15*, 613-624.

163. She, W., and Baroux, C. (2015). Chromatin dynamics in pollen mother cells underpin a common scenario at the somatic-to-reproductive fate transition of both the male and female lineages in Arabidopsis. Front Plant Sci *6*, 294.

164. She, W., Grimanelli, D., Rutowicz, K., Whitehead, M.W., Puzio, M., Kotlinski, M., Jerzmanowski, A., and Baroux, C. (2013). Chromatin reprogramming during the somatic-to-reproductive cell fate transition in plants. Development *140*, 4008-4019.

165. Ravi, M., Shibata, F., Ramahi, J.S., Nagaki, K., Chen, C., Murata, M., and Chan, S.W. (2011). Meiosis-specific loading of the centromere-specific histone CENH3 in Arabidopsis thaliana. PLoS Genet *7*, e1002121.

166. Hsieh, P.H., He, S., Buttress, T., Gao, H., Couchman, M., Fischer, R.L., Zilberman, D., and Feng, X. (2016). Arabidopsis male sexual lineage exhibits more robust maintenance of CG methylation than somatic tissues. Proc Natl Acad Sci U S A *113*, 15132-15137.

167. Komiya, R., Ohyanagi, H., Niihama, M., Watanabe, T., Nakano, M., Kurata, N., and Nonomura, K. (2014). Rice germline-specific Argonaute MEL1 protein binds to phasiRNAs generated from more than 700 lincRNAs. Plant J. *78*, 385-397.

168. Nonomura, K., Morohoshi, A., Nakano, M., Eiguchi, M., Miyao, A., Hirochika, H., and Kurata, N. (2007). A germ cell specific gene of the ARGONAUTE family is essential for the progression of premeiotic mitosis and meiosis during sporogenesis in rice. Plant Cell *19*, 2583-2594.

169. Fei, Q., Yang, L., Liang, W., Zhang, D., and Meyers, B.C. (2016). Dynamic changes of small RNAs in rice spikelet development reveal specialized reproductive phasiRNA pathways. J Exp Bot *67*, 6037-6049.

170. Patel, P., Mathioni, S., Kakrana, A., Shatkay, H., and Meyers, B.C. (2018). Reproductive phasiRNAs in grasses are compositionally distinct from other classes of small RNAs. New Phytol *advance online publication*.

171. Feng, X., Zilberman, D., and Dickinson, H. (2013). A conversation across generations: soma-germ cell crosstalk in plants. Dev Cell *24*, 215-225.

172. Ito, H., Gaubert, H., Bucher, E., Mirouze, M., Vaillant, I., and Paszkowski, J. (2011). An siRNA pathway prevents transgenerational retrotransposition in plants subjected to stress. Nature *472*, 115-119.

173. Iwasaki, Y.W., Siomi, M.C., and Siomi, H. (2015). PIWI-Interacting RNA: Its Biogenesis and Functions. Annu Rev Biochem *84*, 405-433.

174. Andersen, P.R., Tirian, L., Vunjak, M., and Brennecke, J. (2017). A heterochromatin-dependent transcription machinery drives piRNA expression. Nature *549*, 54-59.

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

176. Houben, A., Kumke, K., Nagaki, K., and Hause, G. (2011). CENH3 distribution and differential chromatin modifications during pollen development in rye (Secale cereale L.). Chromosome Res *19*, 471-480.

177. 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.

178. Schoft, V.K., Chumak, N., Mosiolek, M., Slusarz, L., Komnenovic, V., Brownfield, L., Twell, D., Kakutani, T., and Tamaru, H. (2009). Induction of RNA-directed DNA methylation upon decondensation of constitutive heterochromatin. EMBO Rep *10*, 1015-1021.

179. Park, K., Kim, M.Y., Vickers, M., Park, J.S., Hyun, Y., Okamoto, T., Zilberman, D., Fischer, R.L., Feng, X., Choi, Y., et al. (2016). DNA demethylation is initiated in the central cells of Arabidopsis and rice. Proc Natl Acad Sci U S A *113*, 15138-15143.

180. Frost, J.M., Kim, M.Y., Park, G.T., Hsieh, P.H., Nakamura, M., Lin, S.J.H., Yoo, H., Choi, J., Ikeda, Y., Kinoshita, T., et al. (2018). FACT complex is required for DNA demethylation at heterochromatin during reproduction in Arabidopsis. Proc Natl Acad Sci U S A *115*, E4720-E4729.

181. Barckmann, B., Pierson, S., Dufourt, J., Papin, C., Armenise, C., Port, F., Grentzinger, T., Chambeyron, S., Baronian, G., Desvignes, J.P., et al. (2015). Aubergine iCLIP Reveals piRNA-Dependent Decay of mRNAs Involved in Germ Cell Development in the Early Embryo. Cell Rep *12*, 1205-1216.

182. Cao, H., Li, X.Y., Wang, Z., Ding, M., Sun, Y.Z., Dong, F.Q., Chen, F.Y., Liu, L.A., Doughty, J., Li, Y., et al. (2015). Histone H2B Monoubiquitination Mediated by HISTONE MONOUBIQUITINATION1 and HISTONE MONOUBIQUITINATION2 Is Involved in Anther Development by Regulating Tapetum Degradation-Related Genes in Rice. Plant Physiology *168*, 1389-U1514.

183. Solis, M.T., Chakrabarti, N., Corredor, E., Cortes-Eslava, J., Rodriguez-Serrano, M., Biggiogera, M., Risueno, M.C., and Testillano, P.S. (2014). Epigenetic changes accompany developmental programmed cell death in tapetum cells. Plant Cell Physiol *55*, 16-29.

184. Reimegard, J., Kundu, S., Pendle, A., Irish, V.F., Shaw, P., Nakayama, N., Sundstrom, J.F., and Emanuelsson, O. (2017). Genome-wide identification of physically clustered genes suggests chromatin-level co-regulation in male reproductive development in Arabidopsis thaliana. Nucleic Acids Res *45*, 3253-3265.

185. Suwabe, K., Suzuki, G., Takahashi, H., Shiono, K., Endo, M., Yano, K., Fujita, M., Masuko, H., Saito, H., Fujioka, T., et al. (2008). Separated transcriptomes of male gametophyte and tapetum in rice: validity of a laser microdissection (LM) microarray. Plant Cell Physiol *49*, 1407-1416.

186. Ma, Y., Kang, J., Wu, J., Zhu, Y., and Wang, X. (2015). Identification of tapetum-specific genes by comparing global gene expression of four different male sterile lines in Brassica oleracea. Plant Mol Biol *87*, 541-554.

187. Deal, R.B., and Henikoff, S. (2011). The INTACT method for cell type-specific gene expression and chromatin profiling in Arabidopsis thaliana. Nat Protoc *6*, 56-68.

188. Picelli, S., Bjorklund, A.K., Faridani, O.R., Sagasser, S., Winberg, G., and Sandberg, R. (2013). Smart-seq2 for sensitive full-length transcriptome profiling in single cells. Nat Methods *10*, 1096-1098.

189. Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., Davis, A.P., Dolinski, K., Dwight, S.S., Eppig, J.T., et al. (2000). Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat Genet *25*, 25-29.

190. Gene-Ontology-Consortium, T. (2017). Expansion of the Gene Ontology knowledgebase and resources. Nucleic Acids Res *45*, D331-d338.

191. Mi, H., Huang, X., Muruganujan, A., Tang, H., Mills, C., Kang, D., and Thomas, P.D. (2017). PANTHER version 11: expanded annotation data from Gene Ontology and Reactome pathways, and data analysis tool enhancements. Nucleic Acids Res *45*, D183-d189.

192. Braam, J. (2005). In touch: plant responses to mechanical stimuli. New Phytol *165*, 373-389.

193. Kim, J.S., Yamaguchi-Shinozaki, K., and Shinozaki, K. (2018). ER-Anchored Transcription Factors bZIP17 and bZIP28 Regulate Root Elongation. Plant Physiol *176*, 2221-2230.

194. Iwata, Y., Ashida, M., Hasegawa, C., Tabara, K., Mishiba, K.I., and Koizumi, N. (2017). Activation of the Arabidopsis membrane-bound transcription factor bZIP28 is mediated by site-2 protease, but not site-1 protease. Plant J. *91*, 408-415.

195. Ma, Z.X., Leng, Y.J., Chen, G.X., Zhou, P.M., Ye, D., and Chen, L.Q. (2015). The THERMOSENSITIVE MALE STERILE 1 Interacts with the BiPs via DnaJ Domain and Stimulates Their ATPase Enzyme Activities in Arabidopsis. PLoS One *10*, e0132500.

196. Zhang, S.S., Yang, H., Ding, L., Song, Z.T., Ma, H., Chang, F., and Liu, J.X. (2017). Tissue-Specific Transcriptomics Reveals an Important Role of the Unfolded Protein Response in Maintaining Fertility upon Heat Stress in Arabidopsis. Plant Cell *29*, 1007-1023.

197. Deppmann, C.D., Acharya, A., Rishi, V., Wobbes, B., Smeekens, S., Taparowsky, E.J., and Vinson, C. (2004). Dimerization specificity of all 67 B-ZIP motifs in Arabidopsis thaliana: a comparison to Homo sapiens B-ZIP motifs. Nucleic Acids Res *32*, 3435-3445.

198. Lee, E., Lucas, J.R., and Sack, F.D. (2014). Deep functional redundancy between FAMA and FOUR LIPS in stomatal development. Plant J. *78*, 555-565.

199. Shi, D., Ren, A., Tang, X., Qi, G., Xu, Z., Chai, G., Hu, R., Zhou, G., and Kong, Y. (2018). *MYB52* Negatively Regulates Pectin Demethylesterification in Seed Coat Mucilage. Plant Physiol *176*, 2737-2749.

200. Cassan-Wang, H., Goue, N., Saidi, M.N., Legay, S., Sivadon, P., Goffner, D., and Grima-Pettenati, J. (2013). Identification of novel transcription factors regulating secondary cell wall formation in Arabidopsis. Front Plant Sci *4*, 189.

201. Yu, H.J., Hogan, P., and Sundaresan, V. (2005). Analysis of the female gametophyte transcriptome of Arabidopsis by comparative expression profiling. Plant Physiol *139*, 1853-1869.

202. Shibata, M., Breuer, C., Kawamura, A., Clark, N.M., Rymen, B., Braidwood, L., Morohashi, K., Busch, W., Benfey, P.N., Sozzani, R., et al. (2018). GTL1 and DF1 regulate root hair growth through transcriptional repression of ROOT HAIR DEFECTIVE 6-LIKE 4 in Arabidopsis. Development *145*.

203. Voiniciuc, C., Yang, B., Schmidt, M.H.W., Gunl, M., and Usadel, B. (2015). Starting to Gel: How Arabidopsis Seed Coat Epidermal Cells Produce Specialized Secondary Cell Walls. Int J Mol Sci *16*, 3452-3473.

204. Zhu, J., Zhang, G., Chang, Y., Li, X., Yang, J., Huang, X., Yu, Q., Chen, H., Wu, T., and Yang, Z. (2010). AtMYB103 is a crucial regulator of several pathways affecting Arabidopsis anther development. Sci China Life Sci *53*, 1112-1122.

205. Grant, E.H., Fujino, T., Beers, E.P., and Brunner, A.M. (2010). Characterization of NAC domain transcription factors implicated in control of vascular cell differentiation in Arabidopsis and Populus. Planta *232*, 337-352.

206. Zhong, R., Lee, C., Zhou, J., McCarthy, R.L., and Ye, Z.H. (2008). A battery of transcription factors involved in the regulation of secondary cell wall biosynthesis in Arabidopsis. Plant Cell *20*, 2763-2782.

207. Groner, W.D., Christy, M.E., Kreiner, C.M., and Liljegren, S.J. (2016). Allele-Specific Interactions between CAST AWAY and NEVERSHED Control Abscission in Arabidopsis Flowers. Front Plant Sci *7*, 1588.

208. Patharkar, O.R., and Walker, J.C. (2016). Core Mechanisms Regulating Developmentally Timed and Environmentally Triggered Abscission. Plant Physiol *172*, 510-520.

209. Liu, B., Butenko, M.A., Shi, C.L., Bolivar, J.L., Winge, P., Stenvik, G.E., Vie, A.K., Leslie, M.E., Brembu, T., Kristiansen, W., et al. (2013). NEVERSHED and INFLORESCENCE DEFICIENT IN ABSCISSION are differentially required for cell expansion and cell separation during floral organ abscission in Arabidopsis thaliana. J Exp Bot *64*, 5345-5357.

210. Klepikova, A.V., Kasianov, A.S., Gerasimov, E.S., Logacheva, M.D., and Penin, A.A. (2016). A high resolution map of the Arabidopsis thaliana developmental transcriptome based on RNA-seq profiling. Plant J. *88*, 1058-1070.

211. Goff, L., Trapnell, C., and Kelley, D. (2013). cummeRbund: Analysis, exploration, manipulation, and visualization of Cufflinks high-throughput sequencing data. R package version *2*.

212. Beers, E.P., Jones, A.M., and Dickerman, A.W. (2004). The S8 serine, C1A cysteine and A1 aspartic protease families in Arabidopsis. Phytochemistry *65*, 43-58.

213. Alexander, M.P. (1969). Differential staining of aborted and nonaborted pollen. Stain Technol *44*, 117-122.

214. Matsumura, Y., Iwakawa, H., Machida, Y., and Machida, C. (2009). Characterization of genes in the ASYMMETRIC LEAVES2/LATERAL ORGAN BOUNDARIES (AS2/LOB) family in Arabidopsis thaliana, and functional and molecular comparisons between AS2 and other family members. Plant J. *58*, 525-537.

215. Ayadi, M., Delaporte, V., Li, Y.-F., and Zhou, D.-X. (2004). Analysis of GT-3a identifies a distinct subgroup of trihelix DNA-binding transcription factors in Arabidopsis. FEBS Letters *562*, 147-154.

216. Nagata, T., Niyada, E., Fujimoto, N., Nagasaki, Y., Noto, K., Miyanoiri, Y., Murata, J., Hiratsuka, K., and Katahira, M. (2010). Solution structures of the trihelix DNA-binding domains of the wild-type and a phosphomimetic mutant of Arabidopsis GT-1: mechanism for an increase in DNA-binding affinity through phosphorylation. Proteins *78*, 3033-3047.

217. Licausi, F., Weits, D.A., Pant, B.D., Scheible, W.R., Geigenberger, P., and van Dongen, J.T. (2011). Hypoxia responsive gene expression is mediated by various subsets of transcription factors and miRNAs that are determined by the actual oxygen availability. New Phytol *190*, 442-456.

218. O'Malley, R.C., Huang, S.C., Song, L., Lewsey, M.G., Bartlett, A., Nery, J.R., Galli, M., Gallavotti, A., and Ecker, J.R. (2016). Cistrome and Epicistrome Features Shape the Regulatory DNA Landscape. Cell *165*, 1280-1292.

219. Doblas, V.G., Smakowska-Luzan, E., Fujita, S., Alassimone, J., Barberon, M., Madalinski, M., Belkhadir, Y., and Geldner, N. (2017). Root diffusion barrier control by a vasculature-derived peptide binding to the SGN3 receptor. Science *355*, 280-284.

220. Yadav, V., Molina, I., Ranathunge, K., Castillo, I.Q., Rothstein, S.J., and Reed, J.W. (2014). ABCG transporters are required for suberin and pollen wall extracellular barriers in Arabidopsis. Plant Cell *26*, 3569-3588.

221. Yim, S., Khare, D., Kang, J., Hwang, J.U., Liang, W., Martinoia, E., Zhang, D., Kang, B., and Lee, Y. (2016). Postmeiotic development of pollen surface layers requires two Arabidopsis ABCG-type transporters. Plant Cell Rep *35*, 1863-1873.

222. Jin, H., Song, Z., and Nikolau, B.J. (2012). Reverse genetic characterization of two paralogous acetoacetyl CoA thiolase genes in Arabidopsis reveals their importance in plant growth and development. Plant J. *70*, 1015-1032.

223. Ishiguro, S., Nishimori, Y., Yamada, M., Saito, H., Suzuki, T., Nakagawa, T., Miyake, H., Okada, K., and Nakamura, K. (2010). The Arabidopsis FLAKY POLLEN1 gene encodes a 3-hydroxy-3-methylglutaryl-coenzyme A synthase required for development of tapetum-specific organelles and fertility of pollen grains. Plant Cell Physiol *51*, 896-911.

224. Thompson, E.P., Wilkins, C., Demidchik, V., Davies, J.M., and Glover, B.J. (2010). An Arabidopsis flavonoid transporter is required for anther dehiscence and pollen development. J Exp Bot *61*, 439-451.

225. Bernoux, M., Timmers, T., Jauneau, A., Briere, C., de Wit, P.J., Marco, Y., and Deslandes, L. (2008). RD19, an Arabidopsis cysteine protease required for RRS1-R-mediated resistance, is relocalized to the nucleus by the Ralstonia solanacearum PopP2 effector. Plant Cell *20*, 2252-2264.

226. Coll, N.S., Vercammen, D., Smidler, A., Clover, C., Van Breusegem, F., Dangl, J.L., and Epple, P. (2010). Arabidopsis type I metacaspases control cell death. Science *330*, 1393-1397.

227. Carrie, C., Murcha, M.W., Millar, A.H., Smith, S.M., and Whelan, J. (2007). Nine 3-ketoacyl-CoA thiolases (KATs) and acetoacetyl-CoA thiolases (ACATs) encoded by five genes in Arabidopsis thaliana are targeted either to peroxisomes or cytosol but not to mitochondria. Plant Mol Biol *63*, 97-108.

228. Ream, T.S., Haag, J.R., Wierzbicki, A.T., Nicora, C.D., Norbeck, A.D., Zhu, J.K., Hagen, G., Guilfoyle, T.J., Pasa-Tolic, 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.

229. Pimentel, H., Bray, N.L., Puente, S., Melsted, P., and Pachter, L. (2017). Differential analysis of RNA-seq incorporating quantification uncertainty. Nat Methods *14*, 687-690.

230. Bray, N.L., Pimentel, H., Melsted, P., and Pachter, L. (2016). Near-optimal probabilistic RNA-seq quantification. Nat Biotechnol *34*, 525.

231. Rangwala, S.H., Elumalai, R., Vanier, C., Ozkan, H., Galbraith, D.W., and Richards, E.J. (2006). Meiotically stable natural epialleles of Sadhu, a novel Arabidopsis retroposon. PLoS Genet *2*, e36.

232. Alonso, J.M., Stepanova, A.N., Leisse, T.J., Kim, C.J., Chen, H., Shinn, P., Stevenson, D.K., Zimmerman, J., Barajas, P., Cheuk, R., et al. (2003). Genome-wide insertional mutagenesis of Arabidopsis thaliana. Science *301*, 653-657.

233. Clough, S.J., and Bent, A.F. (1998). Floral dip: a simplified method for Agrobacterium-mediated transformation of Arabidopsis thaliana. The Plant journal : for cell and molecular biology *16*, 735-743.

234. Trapnell, C., Roberts, A., Goff, L., Pertea, G., Kim, D., Kelley, D.R., Pimentel, H., Salzberg, S.L., Rinn, J.L., and Pachter, L. (2012). Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. Nat Protoc *7*, 562-578.

235. Gursanscky, N.R., Jouannet, V., Grunwald, K., Sanchez, P., Laaber-Schwarz, M., and Greb, T. (2016). MOL1 is required for cambium homeostasis in Arabidopsis. Plant J. *86*, 210-220.

236. Stracke, R., Ishihara, H., Huep, G., Barsch, A., Mehrtens, F., Niehaus, K., and Weisshaar, B. (2007). Differential regulation of closely related R2R3-MYB transcription factors controls flavonol accumulation in different parts of the Arabidopsis thaliana seedling. Plant J. *50*, 660-677.

237. Nodine, M.D., and Bartel, D.P. (2010). MicroRNAs prevent precocious gene expression and enable pattern formation during plant embryogenesis. Genes Dev *24*, 2678-2692.

238. Chang, F., Zhang, Z., Jin, Y., and Ma, H. (2014). Cell Biological Analyses of Anther Morphogenesis and Pollen Viability in Arabidopsis and Rice. In Flower Development: Methods and Protocols, J.L. Riechmann and F. Wellmer, eds. (New York, NY: Springer New York), pp. 203-216.

239. Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B., et al. (2012). Fiji: an open-source platform for biological-image analysis. Nat Methods *9*, 676-682.

240. Rueden, C.T., Schindelin, J., Hiner, M.C., DeZonia, B.E., Walter, A.E., Arena, E.T., and Eliceiri, K.W. (2017). ImageJ2: ImageJ for the next generation of scientific image data. BMC Bioinformatics *18*, 529.

241. Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Amin, N., Schwikowski, B., and Ideker, T. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res *13*, 2498-2504.

242. Wickham., H. (2016). ggplot2: Elegant Graphics for Data Analysis, (New York: Springer-Verlag).

243. Mincarelli, L., Lister, A., Lipscombe, J., and Macaulay, I.C. (2018). Defining Cell Identity with Single-Cell Omics. Proteomics *18*, e1700312.

244. Rosa, S., Duncan, S., and Dean, C. (2016). Mutually exclusive sense–antisense transcription at FLC facilitates environmentally induced gene repression. Nat Commun *7*, 13031.

245. Berry, S., Hartley, M., Olsson, T.S., Dean, C., and Howard, M. (2015). Local chromatin environment of a Polycomb target gene instructs its own epigenetic inheritance. Elife *4*, e07205.

246. Trapnell, C. (2015). Defining cell types and states with single-cell genomics. Genome Res *25*, 1491-1498.

247. Zamani Esteki, M., Dimitriadou, E., Mateiu, L., Melotte, C., Van der Aa, N., Kumar, P., Das, R., Theunis, K., Cheng, J., Legius, E., et al. (2015). Concurrent Whole-Genome Haplotyping and Copy-Number Profiling of Single Cells. The American Journal of Human Genetics *96*, 894-912.

248. Prakadan, S.M., Shalek, A.K., and Weitz, D.A. (2017). Scaling by shrinking: empowering single-cell 'omics' with microfluidic devices. Nat Rev Genet *18*, 345-361.

249. Tang, F., Barbacioru, C., Wang, Y., Nordman, E., Lee, C., Xu, N., Wang, X., Bodeau, J., Tuch, B.B., Siddiqui, A., et al. (2009). mRNA-Seq whole-transcriptome analysis of a single cell. Nat Methods *6*, 377-382.

250. Macosko, E.Z., Basu, A., Satija, R., Nemesh, J., Shekhar, K., Goldman, M., Tirosh, I., Bialas, A.R., Kamitaki, N., Martersteck, E.M., et al. (2015). Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets. Cell *161*, 1202-1214.

251. Shulse, C.N., Cole, B.J., Turco, G.M., Zhu, Y., Brady, S.M., and Dickel, D.E. (2018). High-throughput single-cell transcriptome profiling of plant cell types. bioRxiv.

252. Zheng, G.X., Terry, J.M., Belgrader, P., Ryvkin, P., Bent, Z.W., Wilson, R., Ziraldo, S.B., Wheeler, T.D., McDermott, G.P., Zhu, J., et al. (2017). Massively parallel digital transcriptional profiling of single cells. Nat Commun *8*, 14049.

253. Zambelli, F., Vancampenhout, K., Daneels, D., Brown, D., Mertens, J., Van Dooren, S., Caljon, B., Gianaroli, L., Sermon, K., Voet, T., et al. (2017). Accurate and comprehensive analysis of single nucleotide variants and large deletions of the human mitochondrial genome in DNA and single cells. Eur J Hum Genet *25*, 1229-1236.

254. Destouni, A., Zamani Esteki, M., Catteeuw, M., Tsuiko, O., Dimitriadou, E., Smits, K., Kurg, A., Salumets, A., Van Soom, A., Voet, T., et al. (2016). Zygotes segregate entire parental genomes in distinct blastomere lineages causing cleavage-stage chimerism and mixoploidy. Genome Res *26*, 567-578.

255. Smallwood, S.A., Lee, H.J., Angermueller, C., Krueger, F., Saadeh, H., Peat, J., Andrews, S.R., Stegle, O., Reik, W., and Kelsey, G. (2014). Single-cell genome-wide bisulfite sequencing for assessing epigenetic heterogeneity. Nat Methods *11*, 817-820.

256. Angermueller, C., Lee, H.J., Reik, W., and Stegle, O. (2017). DeepCpG: accurate prediction of single-cell DNA methylation states using deep learning. Genome Biol *18*, 67.

257. Cooper, J., Ding, Y., Song, J.Z., and Zhao, K.J. (2017). Genome-wide mapping of DNase I hypersensitive sites in rare cell populations using single-cell DNase sequencing. Nat Protoc *12*, 2342-2354.

258. Jin, W., Tang, Q., Wan, M., Cui, K., Zhang, Y., Ren, G., Ni, B., Sklar, J., Przytycka, T.M., Childs, R., et al. (2015). Genome-wide detection of DNase I

hypersensitive sites in single cells and FFPE tissue samples. Nature *528*, 142-146.

259. Buenrostro, J.D., Wu, B., Litzenburger, U.M., Ruff, D., Gonzales, M.L., Snyder, M.P., Chang, H.Y., and Greenleaf, W.J. (2015). Single-cell chromatin accessibility reveals principles of regulatory variation. Nature *523*, 486-490.

260. Cusanovich, D.A., Daza, R., Adey, A., Pliner, H.A., Christiansen, L., Gunderson, K.L., Steemers, F.J., Trapnell, C., and Shendure, J. (2015). Multiplex single cell profiling of chromatin accessibility by combinatorial cellular indexing. Science *348*, 910-914.

261. Chappell, L., Russell, A.J.C., and Voet, T. (2018). Single-Cell (Multi)omics Technologies. Annu Rev Genomics Hum Genet *19*, 15-41.

262. Macaulay, I.C., Haerty, W., Kumar, P., Li, Y.I., Hu, T.X., Teng, M.J., Goolam, M., Saurat, N., Coupland, P., Shirley, L.M., et al. (2015). G&T-seq: parallel sequencing of single-cell genomes and transcriptomes. Nat Methods *12*, 519-522.

263. Angermueller, C., Clark, S.J., Lee, H.J., Macaulay, I.C., Teng, M.J., Hu, T.X., Krueger, F., Smallwood, S., Ponting, C.P., Voet, T., et al. (2016). Parallel single-cell sequencing links transcriptional and epigenetic heterogeneity. Nat Methods *13*, 229-232.

264. Macaulay, I.C., Svensson, V., Labalette, C., Ferreira, L., Hamey, F., Voet, T., Teichmann, S.A., and Cvejic, A. (2016). Single-Cell RNA-Sequencing Reveals a Continuous Spectrum of Differentiation in Hematopoietic Cells. Cell Rep *14*, 966-977.

265. Wilson, Nicola K., Kent, David G., Buettner, F., Shehata, M., Macaulay, Iain C., Calero-Nieto, Fernando J., Sánchez Castillo, M., Oedekoven, Caroline A., Diamanti, E., Schulte, R., et al. (2015). Combined Single-Cell Functional and Gene Expression Analysis Resolves Heterogeneity within Stem Cell Populations. Cell Stem Cell *16*, 712-724.

266. Saxena, A., Prasad, M., Gupta, A., Bharill, N., Patel, O.P., Tiwari, A., Er, M.J., Ding, W.P., and Lin, C.T. (2017). A review of clustering techniques and developments. Neurocomputing *267*, 664-681.

267. Trapnell, C., Cacchiarelli, D., Grimsby, J., Pokharel, P., Li, S., Morse, M., Lennon, N.J., Livak, K.J., Mikkelsen, T.S., and Rinn, J.L. (2014). The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. Nat Biotechnol *32*, 381-386.

268. Waddington, C.H. (1957). The strategy of the genes; a discussion of some aspects of theoretical biology, (London: Allen & Unwin).

269. Sinha, R., Stanley, G., Gulati, G.S., Ezran, C., Travaglini, K.J., Wei, E., Chan, C.K.F., Nabhan, A.N., Su, T., Morganti, R.M., et al. (2017). Index Switching Causes "Spreading-Of-Signal" Among Multiplexed Samples In Illumina HiSeq 4000 DNA Sequencing. bioRxiv.

270. Cannoodt, R., Saelens, W., Sichien, D., Tavernier, S., Janssens, S., Guilliams, M., Lambrecht, B.N., De Preter, K., and Saeys, Y. (2016). SCORPIUS improves trajectory inference and identifies novel modules in dendritic cell development. bioRxiv.

271. Pettie, S., and Ramachandran, V. (2002). An optimal minimum spanning tree algorithm. J Acm *49*, 16-34.

272. Schiefthaler, U., Balasubramanian, S., Sieber, P., Chevalier, D., Wisman, E., and Schneitz, K. (1999). Molecular analysis of *NOZZLE*, a gene involved in pattern formation and early sporogenesis during sex organ development in *Arabidopsis thaliana*. Proc Natl Acad Sci U S A *96*, 11664-11669.

273. Kiselev, V.Y., Kirschner, K., Schaub, M.T., Andrews, T., Yiu, A., Chandra, T., Natarajan, K.N., Reik, W., Barahona, M., Green, A.R., et al. (2017). SC3: consensus clustering of single-cell RNA-seq data. Nat Methods *14*, 483-486.

274. Rousseeuw, P.J. (1987). Silhouettes - a Graphical Aid to the Interpretation and Validation of Cluster-Analysis. J Comput Appl Math *20*, 53-65.

275. Hastie, T., and Stuetzle, W. (1989). Principal Curves. J Am Stat Assoc *84*, 502-516.

276. Breiman, L. (2001). Random forests. Machine Learning *45*, 5-32.

277. Chen, W., Yu, X.H., Zhang, K., Shi, J., De Oliveira, S., Schreiber, L., Shanklin, J., and Zhang, D. (2011). Male Sterile2 encodes a plastid-localized fatty acyl carrier protein reductase required for pollen exine development in Arabidopsis. Plant Physiol *157*, 842-853.

278. Sugimoto-Shirasu, K., Stacey, N.J., Corsar, J., Roberts, K., and McCann, M.C. (2002). DNA topoisomerase VI is essential for endoreduplication in Arabidopsis. Curr Biol *12*, 1782-1786.

279. Liu, J., Zhang, Y., Qin, G., Tsuge, T., Sakaguchi, N., Luo, G., Sun, K., Shi, D., Aki, S., Zheng, N., et al. (2008). Targeted degradation of the cyclin-dependent kinase inhibitor ICK4/KRP6 by RING-type E3 ligases is essential for mitotic cell cycle progression during Arabidopsis gametogenesis. Plant Cell *20*, 1538-1554.

280. Wang, S., Lu, J., Song, X.F., Ren, S.C., You, C., Xu, J., Liu, C.M., Ma, H., and Chang, F. (2017). Cytological and Transcriptomic Analyses Reveal Important Roles of CLE19 in Pollen Exine Formation. Plant Physiol *175*, 1186-1202.

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

282. Afgan, E., Baker, D., van den Beek, M., Blankenberg, D., Bouvier, D., Cech, M., Chilton, J., Clements, D., Coraor, N., Eberhard, C., et al. (2016). The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2016 update. Nucleic Acids Res *44*, W3-W10.

283. Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics *29*, 15-21.

284. Liao, Y., Smyth, G.K., and Shi, W. (2014). featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics *30*, 923-930.

285. Supek, F., Bosnjak, M., Skunca, N., and Smuc, T. (2011). REVIGO summarizes and visualizes long lists of gene ontology terms. PLoS One *6*, e21800.

286. Cao, X., and Jacobsen, S.E. (2002). Role of the arabidopsis DRM methyltransferases in de novo DNA methylation and gene silencing. Curr Biol *12*, 1138-1144.

287. Yelina, N.E., Lambing, C., Hardcastle, T.J., Zhao, X., Santos, B., and Henderson, I.R. (2015). DNA methylation epigenetically silences crossover hot spots and controls chromosomal domains of meiotic recombination in Arabidopsis. Genes Dev *29*, 2183-2202.

288. Kim, M.Y., and Zilberman, D. (2014). DNA methylation as a system of plant genomic immunity. Trends Plant Sci *19*, 320-326.

289. Frommer, M., Mcdonald, L.E., Millar, D.S., Collis, C.M., Watt, F., Grigg, G.W., Molloy, P.L., and Paul, C.L. (1992). A Genomic Sequencing Protocol That Yields a Positive Display of 5-Methylcytosine Residues in Individual DNA Strands. Proc Natl Acad Sci U S A *89*, 1827-1831.

290. Zilberman, D. (2017). An evolutionary case for functional gene body methylation in plants and animals. Genome Biol *18*, 87.

291. Zhong, X., Du, J., Hale, C.J., Gallego-Bartolome, J., Feng, S., Vashisht, A.A., Chory, J., Wohlschlegel, J.A., Patel, D.J., and Jacobsen, S.E. (2014). Molecular mechanism of action of plant DRM de novo DNA methyltransferases. Cell *157*, 1050-1060.

292. Cao, X., and Jacobsen, S.E. (2002). Locus-specific control of asymmetric and CpNpG methylation by the *DRM* and *CMT3* methyltransferase genes. Proc Natl Acad Sci U S A *99 Suppl 4*, 16491-16498.

293. Shen, E.Z., Chen, H., Ozturk, A.R., Tu, S., Shirayama, M., Tang, W., Ding, Y.H., Dai, S.Y., Weng, Z., and Mello, C.C. (2018). Identification of piRNA Binding Sites Reveals the Argonaute Regulatory Landscape of the C. elegans Germline. Cell *172*, 937-951 e918.

294. Zhang, D., Tu, S., Stubna, M., Wu, W.-S., Huang, W.-C., Weng, Z., and Lee, H.-C. (2018). The piRNA targeting rules and the resistance to piRNA silencing in endogenous genes. Science *359*, 587-592.

295. Buenrostro, J.D., Wu, B., Chang, H.Y., and Greenleaf, W.J. (2015). ATAC-seq: A Method for Assaying Chromatin Accessibility Genome-Wide. Curr Protoc Mol Biol *109*, 21 29 21-29.

296. Bagijn, M.P., Goldstein, L.D., Sapetschnig, A., Weick, E.M., Bouasker, S., Lehrbach, N.J., Simard, M.J., and Miska, E.A. (2012). Function, targets, and evolution of Caenorhabditis elegans piRNAs. Science *337*, 574-578.

297. Shirayama, M., Seth, M., Lee, H.C., Gu, W., Ishidate, T., Conte, D., Jr., and Mello, C.C. (2012). piRNAs initiate an epigenetic memory of nonself RNA in the C. elegans germline. Cell *150*, 65-77.

298. Krueger, F., and Andrews, S.R. (2011). Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. Bioinformatics *27*, 1571-1572.

299. Coleman-Derr, D., and Zilberman, D. (2012). Deposition of histone variant H2A.Z within gene bodies regulates responsive genes. PLoS Genet *8*, e1002988.

300. Dolzhenko, E., and Smith, A.D. (2014). Using beta-binomial regression for high-precision differential methylation analysis in multifactor whole-genome bisulfite sequencing experiments. BMC Bioinformatics *15*, 215.

301. Langmead, B. (2010). Aligning short sequencing reads with Bowtie. Curr Protoc Bioinformatics *Chapter 11*, Unit 11 17.

302. Nakagawa, T., Kurose, T., Hino, T., Tanaka, K., Kawamukai, M., Niwa, Y., Toyooka, K., Matsuoka, K., Jinbo, T., and Kimura, T. (2007). Development of series of gateway binary vectors, pGWBs, for realizing efficient construction of fusion genes for plant transformation. J Biosci Bioeng *104*, 34-41.

303. Wang, L., Dou, K., Moon, S., Tan, F.J., and Zhang, Z.Z. (2018). Hijacking Oogenesis Enables Massive Propagation of LINE and Retroviral Transposons. Cell *174*, 1082-1094 e1012.

304. Donlin, M.J., Lisch, D., and Freeling, M. (1995). Tissue-specific accumulation of MURB, a protein encoded by MuDR, the autonomous regulator of the Mutator transposable element family. Plant Cell *7*, 1989-2000.

305. Song, X., Li, P., Zhai, J., Zhou, M., Ma, L., Liu, B., Jeong, D.H., Nakano, M., Cao, S., Liu, C., et al. (2012). Roles of DCL4 and DCL3b in rice phased small RNA biogenesis. Plant J. *69*, 462-474.

306. Cai, Q., Qiao, L., Wang, M., He, B., Lin, F.M., Palmquist, J., Huang, S.D., and Jin, H. (2018). Plants send small RNAs in extracellular vesicles to fungal pathogen to silence virulence genes. Science *360*, 1126-1129.

307. Turan, S., Galla, M., Ernst, E., Qiao, J., Voelkel, C., Schiedlmeier, B., Zehe, C., and Bode, J. (2011). Recombinase-mediated cassette exchange (RMCE): traditional concepts and current challenges. J Mol Biol *407*, 193-221.

308. Wedeles, C.J., Wu, M.Z., and Claycomb, J.M. (2013). Protection of germline gene expression by the C. elegans Argonaute CSR-1. Dev Cell *27*, 664-671.

309. Fujii, S., and Takayama, S. (2018). Multilayered dominance hierarchy in plant self-incompatibility. Plant Reprod *31*, 15-19.

310. Burghgraeve, N., Simon, S.A., Barral, S., Fobis-Loisy, I., Holl, A.-C., Ponitzki, C., Schmitt, E., Vekemans, X., and Castric, V. (2018). Base-pairing requirements for small RNA-mediated gene silencing of recessive self-incompatibility alleles in Arabidopsis halleri. bioRxiv.

311. Zhang, W., Kollwig, G., Stecyk, E., Apelt, F., Dirks, R., and Kragler, F. (2014). Graft-transmissible movement of inverted-repeat-induced siRNA signals into flowers. Plant J. *80*, 106-121.