2020

# REGULATORY NETWORKS DURING SEED DEVELOPMENT

Ran Tian
*University of Kentucky*, tianransaber@gmail.com
Digital Object Identifier: https://doi.org/10.13023/etd.2020.316

STUDENT AGREEMENT:

I represent that my thesis or dissertation and abstract are my original work. Proper attribution has been given to all outside sources. I understand that I am solely responsible for obtaining any needed copyright permissions. I have obtained needed written permission statement(s) from the owner(s) of each third-party copyrighted matter to be included in my work, allowing electronic distribution (if such use is not permitted by the fair use doctrine) which will be submitted to UKnowledge as Additional File.

I hereby grant to The University of Kentucky and its agents the irrevocable, non-exclusive, and royalty-free license to archive and make accessible my work in whole or in part in all forms of media, now or hereafter known. I agree that the document mentioned above may be made available immediately for worldwide access unless an embargo applies.

I retain all other ownership rights to the copyright of my work. I also retain the right to use in future works (such as articles or books) all or part of my work. I understand that I am free to register the copyright to my work.

REVIEW, APPROVAL AND ACCEPTANCE

The document mentioned above has been reviewed and accepted by the student's advisor, on behalf of the advisory committee, and by the Director of Graduate Studies (DGS), on behalf of the program; we verify that this is the final, approved version of the student's thesis including all changes required by the advisory committee. The undersigned agree to abide by the statements above.

Ran Tian, Student

Dr. Sharyn E. Perry, Major Professor

Dr. Mark Coyne, Director of Graduate Studies

# REGULATORY NETWORKS DURING SEED DEVELOPMENT

---

DISSERTATION

---

A dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the College of Agriculture, Food and Environment at the University of Kentucky

BY

Ran Tian

Lexington, Kentucky

Director: Dr. Sharyn E. Perry,

Professor of Plant and Soil Sciences

Lexington, Kentucky

2020

ABSTRACT OF DISSERTATION

REGULATORY NETWORKS DURING SEED DEVELOPMENT

Seeds make up approximately 70% of the human diet directly so understanding regulatory mechanisms to generate a seed and the embryo it contains is fundamentally important. The LAFL genes encode transcription factors (TF) that are critical for seed development. Three members of LAFL, LEAFY COTYLEDON2 (LEC2), ABSCISIC ACID3 (ABI3) and FUSCA3 (FUS3), are B3 domain factors that bind DNA motifs called RY motifs. While *LEC2* is expressed earlier in embryo development, *ABI3* and *FUS3* are expressed during later development. All three of these TFs can induce embryo-specific programs after completion of germination to different extents up to and including formation of embryos in a process called somatic embryogenesis (SE). ABI3 also contributes to the abscisic acid (ABA) response. Prior work determined direct and indirect targets of FUS3. Here we report on chromatin immunoprecipitation-tiling array experiments to globally map binding sites for ABI3. We also assessed transcriptomes in response to ABI3 by comparing developing *abi3-5* and wild type seeds and combine this information to ascertain direct and indirect responsive ABI3 target genes. ABI3 can directly induce and repress its target genes' transcript accumulation and some intriguing differences exist in *cis* motifs between these groups of genes. Directly regulated targets reflect ABI3's roles in seed maturation, desiccation tolerance, entry into a quiescent state and longevity. Interestingly, ABI3 directly represses a gene encoding a microRNA (*MIR160B*) that targets *AUXIN RESPONSE FACTOR* (*ARF*)*10* and *ARF16* that are involved in establishment of dormancy. In addition, ABI3, like FUS3, regulates genes encoding *MIR156* but while FUS3 only induces genes encoding this product, ABI3 induces these genes during early stages of seed development, but represses these genes during late development. The interplay between ABI3, the

other *LAFL* genes, and the *VP1/ABI3-LIKE* (*VAL*) genes that are involved in the transition to seedling development are examined and reveal complex interactions controlling development.

ABI3 directly regulates all five *DUF1264* (Domain of Unknown Function 1264) members in Arabidopsis, while two other seed transcription factors FUS3 and AGL15 directly control subsets of genes in this family. Arabidopsis genes designated as *DUF1264s* appear to be expressed specifically within seeds and the encoded protein include a domain of unknown function that is highly conserved in various plants. Here, the direct association of the TFs with these genes and effect on transcript accumulation is verified. Also, higher order mutants were generated. Quadruple *duf1264* mutant shows a reduction in SE compared to wild type control. Pentuple *duf1264* mutant shows a hypersensitive response in seedlings to ABA compared to WT.

In an associated project, an Arabidopsis protein called SIN3A ASSOCIATED POLYPEPTIDE 18 (AtSAP18) was investigated. SAP18 is a transcriptional co-regulator that is a component of histone deacetylase (HDAC) complexes which interacts with a TF of interest in the lab, AGL15 to control embryogenesis. A new phenotype of a loss-of-function mutant *sap18* was documented in that the mutant is hypersensitive to ABA treatment compared to Columbia (Col) wild type (WT), suggesting an important role of SAP18 in modulation of ABA response.

Finally, the global targets of AGL15 were identified by combining previous RNA microarrays and ChIP microarrays with RNA-seq and ChIP-seq in this study. Also, some of these regulatory networks were investigated in the important crop plant, *Glycine max.*

KEY WORDS: ABI3, SAP18, DUF1264, AGL15, Embryogenesis, ABA

Ran Tian

7/15/2020

Date

REGULATORY NETWORKS DURING SEED DEVELOPMENT

BY

Ran Tian

Sharyn E. Perry
Director of Dissertation

Mark Coyne
Director of Graduate Studies

7/15/2020
Date

ACKNOWLEDGEMENTS

I would like to thank my mentor Dr. Sharyn Perry for her advice and guidance over the years.

I would like to thank Drs. Arthur G. Hunt, A. Bruce Downie, Jan Smalle for serving on my graduate committee, and Dr. Lisa J. Vaillancourt for agreeing to be my outside examiner.

I would like to acknowledge all members of the Perry lab, Jeanne Hartman, Dr. Priyanka Paul, Sanjay Joshi, for their advice, friendship, and teamwork. I also want to thank the University of Kentucky, and the Department of Plant and Soil for continuous support my Ph.D. studying.

At the same time, I want to thank the professors and scientists who helped me during my Ph.D. studying, especially Dr. Tomokazu Kawashima, Dr. Luke Moe, Dr. Hongyan Zhu, Dr. Manohar Chakrabarti and Dr. Sitakanta Pattanaik.

I would like to thank all my friends including but not limited to B Kirtley Amos, Sumudu Karunadasa, Timothy Shull, Jia Tan, Xiaocheng Yu, Jinge Liu, Qiulin Qin, Umma Fatema, M. Foteh Ali, M. Foteh Ali, Carol Von Lanken, and Anthony Clark.

I thank my family, especially my wife for believing in me.

Table of Contents

List of Tables

List of Figures

## Chapter 1 Literature Review (Genetic activity during early seed development)

(Parts of this chapter have been submitted to Biochemical Journal as an Invited Review with coauthors Mr. Sanjay Joshi, Dr. Priyanka Paul and Dr. Sharyn Perry. I have noted Sanjay Joshi's portion (part 1.1). Dr. Paul's section is not included in this literature review.)

Seeds are essential for human civilization and so understanding the molecular events underpinning seed development is important. In addition, the approach of somatic embryogenesis (SE) is an important propagation and regeneration strategy to increase desirable genotypes, to develop new genetically modified plants to meet agricultural challenges, and, at a basic science level, to test gene function. Here, we review the current understanding of transcription factors (TFs) involved in zygotic embryonic patterning in seeds, as well as TFs necessary and/or sufficient to drive SE programs. We focus on the model plant *Arabidopsis* for which many tools are available.

### 1.1 Overview of Morphogenesis

(This section was written by the co-author of the submitted paper, Mr. Sanjay Joshi)

The process of flowering plant seed development begins with double fertilization where one of the sperm nuclei fuses with the egg cell nucleus forming a diploid zygote. In most angiosperms, a second sperm nucleus fuses with two polar nuclei of the central cell forming a triploid endosperm which supports and nourishes the developing embryo De Smet, Lau et al. (2010). Some recent reviews on double fertilization include (Bleckmann, Alter et al. 2014, Fatema, Ali et al. 2019, Adhikari, Liu et al. , Sprunck 2020). The embryo and the endosperm grow covered by maternal integument tissues that later form the

seed coat (Coen and Magnani 2018). Thus, the seed consists of three genotypes: the embryo with a 1:1 maternal (M): paternal (P) ratio, the triploid endosperm with a 2M:1P ratio, and the seed coat with a diploid 2M genotype.

Upon fertilization, a positive signal from zygote causes the coordinated development of endosperm (Olsen 2004, Nowack, Grini et al. 2006). There have been various studies showing endosperm development is critical for embryo viability (Chaudhury, Ming et al. 1997, Pignocchi, Minns et al. 2009, Lafon-Placette and Köhler 2014). In the majority of angiosperms, the triploid structure of endosperm initially undergoes free nuclear division without cytokinesis forming a syncytium (Li and Berger 2012). Later, the endosperm cellularizes (Boisnard-Lorig, Colon-Carmona et al. 2001, Hehenberger, Kradolfer et al. 2012, Li and Berger 2012). The endosperm is largely transient and is mostly consumed by the embryo in some species, such as *Arabidopsis*, whereas in other plants (e.g., *Zea mays*) the endosperm is persistent and also supports the young seedling (Li and Berger 2012). This review does not address endosperm development. However, for further details on endosperm, the reader may refer to reviews (Olsen 2001, Berger 2003, Li and Berger 2012, Lafon-Placette and Köhler 2014). For some interesting recent literature on endosperm-embryo communication, please see recent reviews (Figueiredo and Kohler 2016, Robert 2019). Imprinting is also a fascinating phenomenon not covered in this review, but some recent reviews include (Han, Bartels et al. 2019, Batista and Kohler 2020).

Embryogenesis is generally considered to occur in two phases: morphogenesis and maturation. Morphogenesis is the phase of embryogenesis during which the plant body

is established, starting with the zygote (De Smet, Lau et al. 2010, Colette, Lu et al. 2015). The primary meristems that give rise to the epidermis, vascular system and ground tissue are established and are called the protoderm, procambium or provascular meristem, and ground meristem, respectively. The shoot and root apical meristems that will perpetuate these primary meristems after completion of germination are also founded.

Cellular polarity is an important mechanism involved in many aspects of plant development, including early embryogenesis. Plant cells cannot generally move due to cell walls, so cellular migration, as found in animals, does not contribute to development. However, many plant cells are in symplastic communication via “channels” called plasmodesmata (PD) that are regulated pathways connecting the plasma membrane, endoplasmic reticulum and cytoplasm cell-to-cell. PD can selectively allow transport of relatively large molecules including transcription factors (TFs) between cells. Some recent reviews include (Otero, Helariutta et al. 2016, Han, Huang et al. 2019).

In many species, such as *Arabidopsis* and *Nicotiana tabacum* (tobacco), angiosperm egg cells are polarized with most of the cytoplasm and nucleus toward the chalazal (apical) end of the ovule (also a polar structure) and vacuoles at the micropylar (basal) end. Generally, the zygote elongates augmenting the asymmetry. However, a transient more symmetrical stage often occurs after fusion with the sperm cell, where the nucleus is located more centrally, and the large vacuole divides into many small, more evenly distributed vacuoles within the cell ((Faure, Rotman et al. 2002, Kimata, Higaki et al. 2016) reviewed in (Palovaara, de Zeeuw et al. 2016)). Repolarization of the zygote involves elongation, nuclear migration towards the chalazal end and formation of a large

vacuole at the micropylar end. The zygote elongates approximately three-fold in *Arabidopsis* (Souter and Lindsey 2000, Zhao, Begcy et al. 2017).

In almost all flowering plant species the zygote undergoes transverse division that may be symmetrical (e.g. *Triticum aestivum*) (Sivaramakrishna 1978, Kumlehn, Lorz et al. 1998), but is more generally asymmetrical (e.g. *Arabidopsis*) (Sivaramakrishna 1978, Mansfield and Briarty 1991). This division leads to an apical daughter cell (AC) which always provides cells that contribute to the embryo proper (EP; the part of the embryo that upon completion of germination develops into the seedling) and basal daughter cell (BC) forming the suspensor and at least part of the root. (Sivaramakrishna 1978, Mansfield and Briarty 1991, West and Harada 1993, Goldberg, De Paiva et al. 1994, Vernoud, Hajduch et al. 2005). There is diversity in contributions of the AC and BC to embryo proper and suspensor (for review (Wang, Chen et al. 2020)). The suspensor is a transient structure that provides growth factors and nutrients to the developing embryo. Interestingly, suspensor cells have greater developmental potential and can express embryo proper programs, up to forming viable embryos when there are defects in the embryo proper (reviewed in (Jacob and Brian)). Morphogenesis for a dicot (*Arabidopsis*) and a monocot (*Zea mays*) are outlined below.

*Arabidopsis* - Many embryogenesis/seed development studies have been focused on the model plant *Arabidopsis* because of abundant molecular and genetic tools available (Gepstein and Horwitz 1995). In addition, morphogenesis consists of a simple and predictable pattern which is invariant (Wendrich and Weijers 2013, Colette, Lu et al. 2015) and visualization of development is relatively easy (Schwartz, Yeung et al. 1994,

Feng and Ma 2017). Not all dicots show this type of predictable pattern of divisions; for example, *Gossypium hirsutum* (cotton) has more random early divisions after the first asymmetric division, but organizes a recognizable protoderm by early globular stage (Reeves 1935), as well as the other primary and apical meristems later in morphogenesis.

In *Arabidopsis*, after the first division, the smaller AC generates a spherical pro-embryo by undergoing three rounds of cell division; first two rounds of longitudinal divisions at right angles to form four cells of equal size, followed by one transverse division that gives two tiers, producing eight-cell stage also called an octant stage embryo (Figure 1.1) (West and Harada 1993, Jürgens 2001, Yoshida, de Reuille et al. 2014, Colette, Lu et al. 2015, Palovaara, de Zeeuw et al. 2016). At the octant stage, the apical-basal axis of the embryo can be distinguished as the apical/upper tier embryo domain (UT), which generates the shoot apical meristem and most of the cotyledons, and the basal/lower tier embryo domain (LT), that will contribute to the hypocotyl, root apical meristem, and parts of the cotyledons. The BC divides transversely generating a filament of seven to nine cells by the globular stage of embryo proper development, giving the extraembryonic suspensor (S), with the exception of the uppermost cell that contributes to the root meristem through the formation of the hypophysis (Figure 1.1). The suspensor connects the pro-embryo to maternal tissue and provides nutrients and growth factors as well as pushing the developing embryo into the nutrient rich endosperm (Laux, Würschum et al. 2004, Lau, Slane et al. 2012, Wendrich and Weijers 2013).

The upper and lower-tier cells undergo a tangential (also called periclinal that means the new cell wall is parallel to the surface) division, giving rise to sixteen cells with

eight inner cells and eight outer cells. This stage is also called the dermatogen stage, because the outermost cells are protodermal cells that form epidermal tissue by anticlinal (perpendicular to the organ surface) divisions (Figure 1.1). Further divisions form the globular stages (32 to 64 cells) where the ground meristem and provascular/procambial meristems are specified that give rise to the ground and vascular tissues, establishing the radial axis (Jürgens 2001, Laux, Würschum et al. 2004, Chandler, Nardmann et al. 2008, Wendrich and Weijers 2013). Also, during this stage, the hypophysis, derived from the topmost cell of the basal cell lineage, divides asymmetrically to form the precursor of the quiescent center (QC), which is a smaller lens-shaped cell and another larger basal cell which is a precursor of the distal stem cells of the root meristem, the columella (Jürgens 2001, Laux, Würschum et al. 2004, Chandler, Nardmann et al. 2008, Wendrich and Weijers 2013).

At the end of the late globular stage, cells at the top peripheral regions begin dividing more rapidly to form the two cotyledons. Because this is a transition from a radially symmetrical globular stage embryo to a bilaterally symmetric heart stage embryo, this stage is sometimes referred to as “transition or triangular stage” (not shown in Figure 1.1). The embryo greens during this stage and vascular differentiation (xylem and phloem) becomes apparent. Continued cell division and expansion generate later heart and then torpedo stages, and the root apical meristem (RAM) is organized. The shoot apical meristem (SAM) that will perpetuate the above ground primary meristems after completion of germination becomes anatomically apparent during torpedo stages as a population of smaller cells. The suspensor degenerates during later morphogenesis

(Bozhkov, Filonova et al. 2005) and further expansion causes the embryo to bend to fit within the seed coat (bent cotyledon and mature stage; Figure 1.1).

Overlapping with and continuing after morphogenesis, maturation programs are active to accumulate storage reserve materials in the seeds. In *Arabidopsis*, this is primarily in the cotyledons during later development because the endosperm is a transient structure that is nearly used up by the mature stage. Additionally, desiccation tolerance is acquired, dormancy programs may be established, and desiccation occurs (Thomas 1993, West and Harada 1993, Laux and Jurgens 1997, Wendrich and Weijers 2013). Some key molecular players in the earlier stages of morphogenesis will be described below. Excellent reviews for the maturation processes include (Gutierrez, Wuytswinkel et al. 2007, Holdsworth, Bentsink et al. 2008, Leprince, Pellizzaro et al. 2017).

## 1.2 Arabidopsis Transcription Factors that Promote Embryo Identity

### 1.2.1 TFs necessary and sufficient for embryogenesis: The LAFLs

While forward genetic screens to identify embryo defective mutants generally result in the identification of non-redundant genes essential for viability rather than for embryo development or designation of domains within the embryo specifically (Jürgens 2001, Meinke 2020), some embryo defective mutants have lesions in genes that are more specific for seed development. These include genes encoding four central regulators LEAFY COTYLEDON1 (LEC1), ABSCISIC ACID INSENSITIVE3 (ABI3), FUSCA3 (FUS3) and LEAFY COTYLEDON2 (LEC2), otherwise referred to as the LAFL factors, that are key to embryo development. These genes are necessary and sufficient for embryo processes and show the highest levels of transcript accumulation in the developing embryo and

endosperm ((Le, Cheng et al. 2010) and *Arabidopsis* eFP Browser (Winter, Vinegar et al. 2007)), although they have also been reported to have developmental roles after completion of germination (Chiu, Nahal et al. 2012, Junker and Baumlein 2012, Junker, Monke et al. 2012, Huang, Hu et al. 2015, Bedi, Sengupta et al. 2016, Tang, Zhou et al. 2017). To be necessary, a loss-of-function of the factor would result in an embryo-defective phenotype (whether specific to embryogenesis or not). Sufficiency is tested by ectopic expression of the factor and assessment of whether it can drive embryo-specific programs ectopically. The LAFL factors are among those necessary and sufficient (to different extents; discussed below) to drive embryo processes. Other TFs may be sufficient to drive embryo programs, but not necessary, perhaps due to redundancy. Others may show embryo deficient phenotypes, indicating necessity, but are insufficient to drive embryo programs, often because they are needed for cell viability rather than embryo specification, but also possibly due to interacting factors missing in the ectopically expressed domains. The LAFLs and other factors interact in a complex network to confer embryo identity. While putative orthologs of LEC1 and ABI3, and perhaps LEC2, appear to have arisen prior to the "invention" of the seed, FUS3 orthologs have only been found in seed plants to date (Han, Li et al. 2017) and reviewed in (Boulard, Fatihi et al. 2017). Roles for putative orthologs in non-seed plants include desiccation tolerance, and many aspects of LAFL-loss-of-function phenotypes indicate that they may have been recruited to control processes during the maturation phase and establish the quiescent state of the mature seed.

LEAFY COTYLEDON 1 (LEC1, At1g21970) is a NF-YB/HAP3 subunit of a CCAAT box binding factor (CBF) (Lotan, Ohto et al. 1998). NF-YBs form complexes with NF-YA/HAP2 and NF-YC/HAP5 subunits, with the NF-YA subunit primarily responsible for recognition of the DNA *cis* element (the CCAAT box) (reviewed in (Jo, Pelletier et al. 2019)). NF-YB subunits include a broadly conserved histone-like domain (the central B domain) that encompasses residues necessary for complex formation as well as residues contributing to binding DNA (Lee, Fischer et al. 2003). The loss-of-function *lec1* shows defects in embryo morphogenesis and embryonic maturation, including decreased accumulation of storage products and loss of acquisition of desiccation tolerance. When developing embryos are rescued prior to desiccation, the resulting *lec1* homozygous seedlings show leaf-like traits in the cotyledons, including development of trichomes (in *Arabidopsis* trichomes are a leaf, but not cotyledon feature) (Meinke 1992, West and Harada 1993, Meinke, Franzmann et al. 1994). *LEC1* is preferentially expressed in the seeds, beginning early in embryo development and declining during maturation as shown in Figure 1.2A (Lotan, Ohto et al. 1998, Casson, Spencer et al. 2005, Nakabayashi, Okamoto et al. 2005, Schmid, Davison et al. 2005, Klepikova, Kasianov et al. 2016). Early defects include abnormal cell divisions in the suspensor (Lotan, Ohto et al. 1998). The loss-of-function mutant also shows precocious activation of the shoot apical meristem during embryogenesis. When *LEC1* is constitutively expressed using a *35S* promoter (*35S* is a promoter from a virus that promotes broad and generally, but not universally, high level expression in dicots (Dey, Sarkar et al. 2015)), the genetically engineered plants produced somatic embryos on the “seedlings”, which indicates that LEC1 is sufficient to promote

embryo identity (Lotan, Ohto et al. 1998). However, this "sufficiency" has a timeframe. When the *35S:LEC1* transgene included a glucocorticoid receptor (GR) domain that allows the TF to only be transported into the nucleus upon hormone treatment with the synthetic steroid dexamethasone, only early timepoints during or shortly after completion of germination resulted in embryo/embryo-like tissue. By four days after germination, the seedlings did not produce this tissue and instead appeared normal (Junker, Monke et al. 2012). The central B domain of LEC1 protein was confirmed to be required and sufficient for embryogenesis. Moreover, an amino acid residue (aspartic acid at position 55), which specifically exists in LEC1-type B domain proteins (LEC1 and L1L), but not in non-LEC1-type NF-YB subunits that have a lysine in this position, was verified to be necessary for embryogenesis (Lee, Fischer et al. 2003). High throughput analyses to assess the transcriptome in response to *LEC1* overexpression, *lec1* mutant, as well as assessing direct targets using high-throughput chromatin immunoprecipitation (ChIP) experiments have revealed global direct and indirect targets for LEC1 (Junker, Monke et al. 2012, Pelletier, Kwong et al. 2017). These direct targets include the other members of the LAFL group, as well as genes involved in a broad range of processes active during different stages of seed development (Pelletier, Kwong et al. 2017). LEC1 interacts with other proteins such as bZIP67, LEC2, and others to control various programs during seed development (Figure 1.2B and (Pelletier, Kwong et al. 2017, Jo, Pelletier et al. 2019)). LEC1 has been shown to function as a pioneer TF to "reset" expression of *FLOWERING LOCUS C* (*FLC*) during embryogenesis to allow transition to flowering in response to vernalization later in the life cycle (Tao, Shen et al. 2017). Some of the interactions between LEC1 and

other TFs involved in embryogenesis are shown in Figure 1.2B (Lotan, Ohto et al. 1998, Casson, Spencer et al. 2005, Nakabayashi, Okamoto et al. 2005, Schmid, Davison et al. 2005, Klepikova, Kasianov et al. 2016).

LEAFY COTYLEDON 2 (LEC2, At1g28300) contains a B3 domain and is another major TF involved in embryogenesis in plants. Loss-of-function of *LEAFY COTYLEDON*2 (*lec2*) shows a similar phenotype to *lec1* in that both have trichomes on their cotyledons upon embryo rescue, and defects in embryo shape (Meinke, Franzmann et al. 1994). *LEC2* transcript is mainly detected during seed development (Figure 1.2A), but also was detected in other types of tissue in *Arabidopsis* (Stone, Kwong et al. 2001, Casson, Spencer et al. 2005, Nakabayashi, Okamoto et al. 2005, Schmid, Davison et al. 2005, Klepikova, Kasianov et al. 2016). Ectopic *LEC2* expression in vegetative cells could induce a somatic embryo-like seedling phenotype demonstrating sufficiency to drive embryo programs (Stone, Kwong et al. 2001). Also, like LEC1, there is a timeframe for sufficiency. When a *35S:LEC2* transgene is expressed, the "seedlings" produce masses of somatic embryos (Stone, Kwong et al. 2001). But when an inducible form (*35S:LEC2-GR*) is used, while seedlings treated with dexamethasone activate embryo programs and leaves take on embryo-like features, they do not produce somatic embryos (Braybrook, Stone et al. 2006, Stone, Braybrook et al. 2008, Feeney, Frigerio et al. 2013). The inducible form of LEC2 was used to globally identify putative direct targets revealing the regulation of *YUCCA* genes involved in auxin biosynthesis. (Braybrook and Harada 2008, Stone, Braybrook et al. 2008, Wojcikowska, Jaskola et al. 2013, Wojcikowska and Gaj 2015). Like LEC1, LEC2 (and FUS3) are involved in the epigenetic resetting of *FLC* (Tao, Hu et al. 2019).

FUSCA3 (FUS3, At3g26790) is a B3 domain TF. Loss-of-function of *fus3* seeds are desiccation intolerant, but like *lec1* and *lec2*, can be rescued by culturing prior to drying, upon which the cotyledons show "leafy" traits (trichomes), but these trichomes were much less prevalent compared to *lec1* or *lec2* mutant plants and the embryos have a more normal bent cotyledon shape (Keith, Kraml et al. 1994, Meinke, Franzmann et al. 1994). Unlike LEC1 and LEC2 overexpression that produces somatic embryos on the seedlings, the FUS3 ectopic phenotype was milder and produced cotyledon-like leaves. (Gazzarrini, Tsuchiya et al. 2004). Thus, FUS3 is a key regulator during seed development as shown by being necessary and sufficient to specify embryo programs. *FUS3* transcript is primarily expressed in seeds but is also detected at low levels in tissues after the completion of germination (Luerssen, Kirik et al. 1998). *FUS3* transcript accumulation begins at the 8-cell embryo stage, and continues to be expressed relatively late during embryo development compared with *LEC1/LEC2* (Figure 1.2A and (Baumlein, Misera et al. 1994, Tsuchiya, Nambara et al. 2004, Casson, Spencer et al. 2005, Nakabayashi, Okamoto et al. 2005, Schmid, Davison et al. 2005, To, Valon et al. 2006, Willmann, Mehalick et al. 2011, Klepikova, Kasianov et al. 2016, Roscoe, Vaissayre et al. 2019)). FUS3 is positively regulated by the plant hormone abscisic acid (ABA) and negatively regulated by gibberellic acid (GA) (Lu, dela Paz et al. 2010, Chiu, Saleh et al. 2016). Whole-genome analysis of direct and indirect targets of FUS3 has been performed (Yamamoto, Kagaya et al. 2010, Wang and Perry 2013).

ABSCISIC ACID-INDEPENDENT3 (ABI3, At3g24650) is the third B3 domain TF of the LAFL genes. Like *lec1*, *lec2*, and *fus3* mutant, *abi3* seeds are desiccation intolerant but

homozygous plants could be generated by culture rescuing before the seed dries. Unlike the other LAFL family members, *abi3* did not show abnormal leafy cotyledons, but did show reduced seed storage protein accumulation. ABI3 is also involved in ABA perception and response (Giraudat, Hauge et al. 1992, Nambara, Naito et al. 1992). *ABI3* transcripts accumulate primarily in seeds with a transient expression detected after seeds complete germination. Ectopic expression of *ABI3* did not produce embryos on vegetative tissues, but some seed-specific storage protein mRNA was found in seedlings indicating the ability to ectopically activate some embryo programs (Parcy, Valon et al. 1994). *ABI3* is expressed at higher levels during the late stages of embryo development (maturation phase) (Parcy, Valon et al. 1994, Casson, Spencer et al. 2005, Nakabayashi, Okamoto et al. 2005, Schmid, Davison et al. 2005, To, Valon et al. 2006, Klepikova, Kasianov et al. 2016). Genome-wide targets of ABI3 regulation have been identified (Mönke, Seifert et al. 2012, Tian, Wang et al. 2020).

*LEAFY COTYLEDON1-LIKE* (*L1L, At5g47670*), was isolated after the original LAFLs, but is the closest related gene to *LEC1* with 83% sequence identity and encodes a LEC-1 type NF-YB subunit. *L1L is* primarily expressed in developing seeds and transcript accumulation is highest during maturation (Figure 1.2A). The *l1l* RNAi suppression lines showed an incompletely penetrant and range of phenotypes varying from arrest at the globular stage with abnormal suspensor divisions to underdeveloped cotyledons. Although somatic embryos were not produced on any vegetative tissues examined in *35S:L1L*, the young seedlings did ectopically express embryo-specific programs (e.g. accumulation of storage products). In addition, the over-expression of *LIL* could suppress

the desiccation intolerant phenotype of the *lec1* mutant as could *LIL* driven by the *LEC1* regulatory sequences (Kwong, Bui et al. 2003).

### 1.2.2 Genes that are sufficient for embryo identity/somatic embryogenesis (SE)

In addition to the key LAFL factors described above, a number of other genes were found to be sufficient to confer embryo identity when ectopically expressed but were not required (necessary) for normal zygotic embryo development. This may be due to redundant gene function(s) in the embryo (Cutler and McCourt 2005). Some of these genes, can, like *LEC1* and *LEC2* promote SE on various parts of the seedlings. Others only promote SE on particular tissues (e.g., roots; for example, *WUS*) or in particular culture conditions (e.g., *AGL15*, *AGL18*).

*AGAMOUS-Like15* (*AGL15*, At5g13790) and *AGAMOUS-Like18* (*AGL18*, At3g57390) encode MADS domain transcription factors that are preferentially, but not exclusively, expressed during seed development (Heck, Perry et al. 1995, Rounsley, Ditta et al. 1995). Over-expression of either of these factors dramatically promoted secondary embryo production from zygote embryo explants in the absence of exogenous hormones (Harding, Tang et al. 2003, Adamczyk, Lehti-Shiu et al. 2007). Subculturing allowed continued somatic embryo development in the absence of any exogenously added hormones for a long period time with the oldest *35S:AGL15* cultures being nearly 24 years old (Harding, Tang et al. 2003, Adamczyk, Lehti-Shiu et al. 2007, Zheng, Ren et al. 2009). Additionally, both AGL15 and AGL18 could promote SE from the shoot apical meristem when mature seeds were allowed to complete germination in liquid medium containing the synthetic auxin 2,4-D ((Harding, Tang et al. 2003) and Paul and Perry, unpublished) and these

proteins can interact (Serivichyaswat, Ryu et al. 2015). Direct and indirect targets for AGL15 have been elucidated and include *LEC2, ABI3* and *FUS3* (Figure 1.2B) and (Zheng, Ren et al. 2009). Interestingly, VAL1 directly represses *AGL15* during the transition from seed to seedling development (Chen, Veerappan et al. 2018).

While *SOMATIC EMBRYOGENESIS RECEPTOR KINASE 1* (*AtSERK1, At1g71830*) does not encode a TF, but rather encodes a leucine-rich repeat (LRR) transmembrane receptor-like kinase (RLK) which is broadly expressed including in developing seeds especially in ovules and embryo, it can also enhance embryonic competence when ectopically expressed (Hecht, Vielle-Calzada et al. 2001). It has been reported to form a complex that includes AGL15 (Karlova, Boeren et al. 2006). Phosphorylation of AGL15 has been reported and is important for function (Patharkar and Walker 2015, Patharkar, Macken et al. 2016).

*WUSCHEL* (*WUS, At2g17950*) also confers embryogenic competence on seedling tissues. For example, overexpression of WUS tissues including seedlings, shoot apical meristem and shoot apical meristem when ectopically expressed. Over-expression of *WUS* promoted somatic embryo formation from root tips of seedlings of *Arabidopsis* (Zuo, Niu et al. 2002, Gallois, Nora et al. 2004, Jha, Ochatt et al. 2020).

Several APETALA2 (AP2) domain TFs (BABY BOOM/PLETHORA4 (BBM/PLT4 At5g17430), EMBRYOMAKER/AINTEGUMENTA-LIKE5/PLT5 (EMK, At5g57390), *PLETHORA2 (PLT2,* At1g51190) and some of the WOUND INDUCED DEDIFFERENTIATION (WIND) factors) promote SE from after completion of germination in *Arabidopsis* when

ectopically expressed (Boutilier, Offringa et al. 2002, Aida, Beis et al. 2004, Su, Zhao et al. 2009, Tsuwamoto, Yokoi et al. 2010, Ikeuchi, Iwase et al. 2015). BBM regulates several of the LAFL genes (Horstman, Li et al. 2017). Other members of the PLT family (*PLT1, 2, 3* and *7*) can also induce SE when ectopically expressed (Horstman, Li et al. 2017).

*PLANT GROWTH ACTIVATOR37* (*PGA37*, At3g27785), encodes a MYB118 transcription factor, and is another important regulator during embryo development. MYB118 is produced in endosperm and embryos. Similar to the *WUS* overexpression phenotype, ectopic expression of *PGA37* by an inducible construct using 17-beta-estradiol promoted SE on root explants. Over-expression of *MYB115* (AT5G40360), a paralogous gene of *PGA37*, showed a similar trend in that ectopic expression could promote SE from root explants of *Arabidopsis* (Wang, Niu et al. 2009).

The *Arabidopsis* genome contains five *RWP-RK DOMAIN-CONTAINING* (*RKD*) genes, and expression of *AtRKD1* (*At1g18790*) and *AtRKD2 (At1g74480*) were detected in reproductive organs, mainly expressed in the egg cell. Single or double mutants of these two genes did not show any phenotype or embryo morphological differences compared with wild type; but over-expression of *AtRKD1* or *AtRKD*2 individually led to the proliferation of tissue expressing egg cell markers (Koszegi, Johnston et al. 2011). *AtRKD4(At5g53040)* is another member of this family that is preferentially expressed in the early embryo proper and suspensor. Deficiencies in RKD4 (also called GROUNDED) causes abnormal zygotic cell elongation and subsequent division, and in general the mutant plants show absence or short main roots. Increased ectopic expression of *GRD/RKD4* resulted in embryo-like structures forming on the roots (Waki, Hiki et al. 2011).

### 1.2.3 Loss-of-function mutants that produce ectopic embryos

A number of other genes promote SE in *Arabidopsis* when present as loss-of-function alleles or transgenics that reduce production of the gene product. These genotypes include *pkl* (Ogas, Cheng et al. 1997, Ogas, Kaufmann et al. 1999), *val1 val2* (Suzuki, Wang et al. 2007, Suzuki and McCarty 2008) (VAL1 is also called HSI2; and VAL2 is also called HSL1), *clf swn* (Chanvivattana, Bishopp et al. 2004), *hda6 hda19* RNAi (Tanaka, Kikuchi et al. 2008), and a double knock-out/down of *AtBMI A/B* (Bratzel, Lopez-Torrejon et al. 2010) or *AtRING1a/b* (Chen, Molitor et al. 2010). These mutants/RNAi plants produce embryos/embryonic calli on aberrant seedlings and may be considered as having defects in the transition from embryo to post-germinative development. Interestingly, this group of genes encode proteins that include domains shown or predicted to be involved in epigenetic control of gene expression and in particular involving repression of gene expression. In cases where it has been investigated, LAFL and other genes mentioned above are derepressed in the seedlings, and in the case of AGL15, VAL1 directly represses gene expression (Figure 1.2B and (Rider, Henderson et al. 2003, Suzuki, Wang et al. 2007, Tanaka, Kikuchi et al. 2008, Bratzel, Lopez-Torrejon et al. 2010, Chen, Molitor et al. 2010, Chen, Veerappan et al. 2018)). VAL1 forms protein interactions with HDA6, while VAL2/HSL1 interacts with HDA19 (Figure 1.2B and (Zhou, Tan et al. 2013, Chhun, Chong et al. 2016). VAL1 and VAL2 have also been reported to interact (Chhun, Chong et al. 2016, Chen, Wang et al.). Other studies found evidence for direct repression of LAFL genes by VAL-HDA complexes, but these studies used *35S* promotor for ChIP and a different stage of seedling development (Zhou, Tan et al. 2013, Chhun, Chong et al.

2016), rather than the native promoter as used in Chen et al., 2018 (Chen, Veerappan et al. 2018) where the association of VAL1 with LAFL regulatory regions was not found. In addition, CLF and SWN that are components of the Polycomb Repressive Complex2 (PRC2) involved in depositing trimethylation on lysine-25 of histone 3 (H2K27me3 – an epigenetic mark associated with repression of expression), may both directly regulate *LEC2, FUS3, ABI3*. *SWN*, but not *CLF*, may directly regulate LEC1 (Shu, Chen et al. 2019).

### 1.2.4 Interaction network among these transcription factors

Genes responsive to accumulation of many of the factors mentioned above have been identified, and in some cases, those that are bound by the TF have been determined (so called direct responsive targets) (Braybrook, Stone et al. 2006, Zheng, Ren et al. 2009, Busch, Miotk et al. 2010, Mönke, Seifert et al. 2012, Wang and Perry 2013, Horstman, Li et al. 2017, Pelletier, Kwong et al. 2017, Tian, Wang et al. 2020). Extensive interaction exists among these genes and the products that they encode, and some direct regulatory interactions are shown in Figure 1.2B. At the bent cotyledon (B-COT) stage during embryo development, LEC1 could directly bind and regulate its own coding region, *LEC2, FUS3, ABI3, AGL15, AtSERK1, L1L, PGA37,* and *AtEMK*. At the same stage, *MYB115* transcripts accumulated more than five-hundred fold when comparing the peripheral endosperm with the distal seed coat (control) (Pelletier, Kwong et al. 2017). This may be a result of indirect regulation because LEC1 was not found to bind to *MYB115* regulatory sequences. In addition, *cis* elements referred to as G-boxes, ABRE-like, CCAAT, RY, and BPC1 are over-represented in target genes bound by LEC1 (Pelletier, Kwong et al. 2017). RY-motifs, that are binding sites for B3 domain TFs, are also over-represented in the gene lists that are

bound by LEC2, FUS3 and ABI3; and G-boxes are also found overrepresented in binding regions associated with FUS3 and ABI3 (Mönke, Seifert et al. 2012, Wang and Perry 2013, Tian, Wang et al. 2020). LEC2 directly regulates *AGL15* (Braybrook, Stone et al. 2006) and indirectly regulates *FUS3* and *ABI3* (To, Valon et al. 2006). Interestingly, LEC2 interacted with FUS3 *in vitro* by using pull-down method and *in vivo* by using BiFC assays (Tang, Zhou et al. 2017). LEC1 and LEC2 can be coimmunoprecipitated in the presence of NF-YC2 and OLE1 promoter (a target DNA sequence) in pull-down experiments (Baud, Kelemen et al. 2016). FUS3-direct target genes include *LEC1, AGL15, ABI3, BBM, L1L,* as well as itself (Wang and Perry 2013). ABI3 could regulate itself and *FUS3* (To, Valon et al. 2006, Tian, Wang et al. 2020). Although WUS could bind to G-box motifs, WUS did not bind to any of the genes mentioned above in shoot apices of *Arabidopsis* (Busch, Miotk et al. 2010) or interact with any of them. This suggests that WUS may participate in embryo development through a different pathway, act at relatively down-stream position in the pathway, or the shoot apex prevents binding to these genes by some undiscovered means. Using microarrays to compare transcriptomes in a *MYB118* over-expression transgenic line with that of the wild type plant showed increased *FUS3* transcript accumulation in response to increased MYB118 (Zhang, Cao et al. 2009). Finally, auxin is an important component of zygotic and somatic embryogenesis, and a number of TFs shown in Figure 1.2, including LEC1, LEC2, BBM, AGL15 and WUS are regulated by auxin and/or directly control auxin biosynthetic and response genes (reviewed in (Wojcik, Wojcikowska et al. 2020)). In addition, microRNAs are involved in SE. As an example, *PLT1* and *PLT2* were

found to be regulated by miRNA396 via GROWTH REGULATING FACTOR (GRF) TFs (Szyrajew, Bielewicz et al. 2017, Szczygiel-Sommer and Gaj 2019).

### 1.2.5 Increased SE by heterologous expression of Arabidopsis genes or expression of native orthologs in other plants

(This and the next section of this chapter were not included in the submitted paper.)

Certainly, some of the genes mentioned above can induce or enhance SE not merely in the model plant *Arabidopsis* but also in many other plants including soybean, cotton, oil palm, and citrus. These experiments provide us with important evidence for understanding the role of these genes and are briefly discussed below.

Overexpression of Arabidopsis *LEC1* (*AtLEC1*) or *AtLEC2* could induce embryo-type structures without need for exogenous auxin in vegetative tissues of tobacco (*Nicotiana tabacum*)(Guo, Liu et al. 2013). A recent report showed over-expression of Cs*FUS*3, a citrus ortholog of *Arabidopsis* FUS3, could promote SE in a SE-recalcitrant callus tissue (Liu, Ge et al. 2018). Ectopic expression of *AGL15* orthologs in some other plants also could promote SE. 35S-promoter-driven AGL15 orthologs in both *Glycine max* (soybean; *GmAGL15*) and the ortholog from *Gossypium hirsutum* (cotton; *GhAGL15*) could enhance somatic embryogenesis (reference?). In particular, SE was greatly increased by *35S:GhAGL15* when auxin was present (reference?). Moreover, the closest paralogs of *AGL15*, called *AGL18*, also its soybean ortholog (*GmAGL18*) could promote SE in *Arabidopsis* and soybean respectively (Thakare, Tang et al. 2008, Yang, Li et al. 2014, Zheng and Perry 2014). There were several reports that showed ectopic expression of

*WUS,* in some cases combined with other factors/hormones, could induce vegetative-to-embryonic transition in *Arabidopsis* (Gallois, Woodward et al. 2002, Zuo, Niu et al. 2002, Gallois, Nora et al. 2004). Heterologous expression of Arabidopsis *WUS* driven by beta-estradiol induced promoters, in the presence of the synthetic cytokinin, 6-benzyladenine (BA), could induce embryo-like structures to form from the root-tip of tobacco (Rashid, Yamaji et al. 2007). After Boutilier's group found over-expression of *Arabidopsis BBM* (*AtBBM*) or Brassica *BBM* (*BnBBM*) induced SE in *Arabidopsis* (Boutilier, Offringa et al. 2002), there were several reports that different BBM orthologs in other plants could also improve SE. Ectopic expression of *AtBBM* or *BnBBM* could not induce any SE in tobacco; but interestingly, somatic embryo formation was observed at the shoot-root junction region (collet) when certain concentrations of benzylaminopurine or zeatin were applied to tobacco four-week seedlings in cultures (Srinivasan, Liu et al. 2007). Mis-expression of either *Arabidopsis* orthologs *BBM* from soybean (*GmBBM*) and from oil palm (*Elaeis guineensis* Jacq; *EgAP2-1*) could also induce SE on cotyledons of *Arabidopsis*, without any exogenous hormone; but SE induction was weaker than for the transgenic *BnBBM Arabidopsis* lines (Morcillo, Gallard et al. 2007, El Ouakfaoui, Schnell et al. 2010). Over expression of another *AtBBM* orthologous gene of *Theobroma cacao* (*TcBBM*), not only induced somatic embryos to form in Arabidopsis but also in the cacao tree, independent of any other inducer (Florez, Erwin et al. 2015). A recent study showed that over-expression of *CcSerk1,* the homolog of *AtSerk1* of *Coffea canephora*, could enhance SE in embryogenesis induction media; about twice the number of somatic embryos were formed on leaf explants of over-expressing *CcSERK1* plants compared with those from

wild type plants (Perez-Pascual, Jimenez-Guillen et al. 2018). *Citrus sinensis L1l* (*CsL1l*) was preferentially expressed in embryogenic callus, somatic embryos and immature seeds of *Citrus sinensis*. The 35S promoter driven *CsL1l* cDNA could induce embryo-like structures to form on the cuttings of epicotyls In *Citrus sinensis* (Zhu, Wang et al. 2014). There were two reports that showed that over-expressing *AtRKD4* induces somatic embryos in *Phalaenopsis* "Sogo Vivien" (an orchid hybrid) protocorm and *Dendrobium 22halaenopsis* orchid protocorm (Mursyanti, Aziz et al. 2015, Setiari, Purwantoro et al. 2018).

### 1.2.6 Expression pattern in early stage of embryo development in Arabidopsis

Approximately 65% of Arabidopsis genes have been detected to be transcribed in the developing embryo, which include the fourteen genes discussed above (*LEC1, LEC2, FUS3, ABI3, AGL15, WUS, BBM, ATSERK1, RKD2, L1L, PGA37/MYB118, MYB115, ATEMK, PLT2*), except *RKD1* and *RKD4* which were not in the DNA microarray database. As discussed, these different genes are distinctively expressed both temporally and spatially during embryo development. Within each genes' transcript accumulation profile, it can be found that *LEC1, LEC2, ABI3, WUS, ATSERK1, L1L, ATEMK*, PLT2 and *RKD2* show highest transcript accumulation at the torpedo stage of development. More specifically, *LEC1* and *AtSERK1* transcripts were detected at the maximum amounts in the apical portion of the torpedo stage embryo. *LEC2, ABI3*, and *L1*L transcript accumulation were the highest in the developing cotyledons at the torpedo stage. *RKD2* and *PLT2* expression levels were detected the highest at the shoot meristem at this stage. *WUS* and *AtSERK1* expressed the highest in the shoot meristem and apical region during the torpedo stage, respectively. *FUS3*, *PGA37*/*MYB118*, and *AGL15* transcription levels were the maximal in the root

portion of the heart stage. *BBM* produced transcripts at the highest level at the basal position at the globular stage. On the contrary, *MYB115* transcripts levels reached the peak at apical portion of the same stage during embryo development (summarized from: The Affymetrix ATH1 array data from (Casson, Spencer et al. 2005) and the images that were generated with the Tissue Specific Embryo Development eFP at bar.utoronto.ca/eplant by (Waese, Fan et al. 2017).

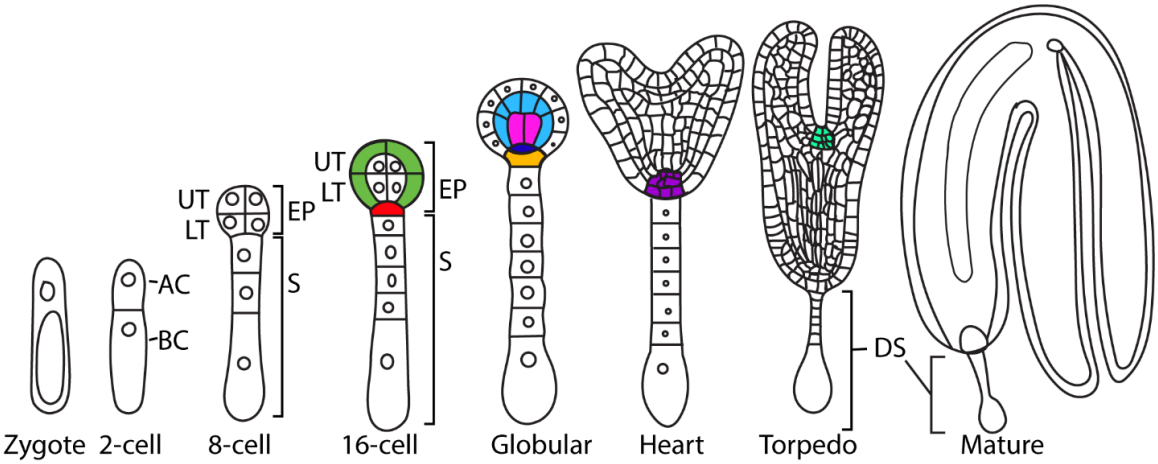



**Figure 1.1**: Different stages of embryo development in Arabidopsis (Dicot). Schematic representation of key embryo stages. AC, apical cell; BC, basal cell; UT upper tier cells; LT lower tier cells; EP, embryo proper; S, suspensor; DS degenerating/degenerated suspensor. Establishment of the: protoderm, green; hypophysis, red; ground meristem, light blue; provascular meristem, pink; lens-shaped cell, dark blue; basal cell from hypophysis, gold; root apical meristem, purple; and shoot apical meristem, mint. No common scales were used in drawings. Some nuclei are not shown as meristems become established, and from heart stage on. Cells are approximate from heart stage on and not shown in the mature embryo. Adapted from (Goldberg, De Paiva et al. 1994, Vernoud, Hajduch et al. 2005). Figure contributed by Ms. Ju-young (Gloria) Yoon.

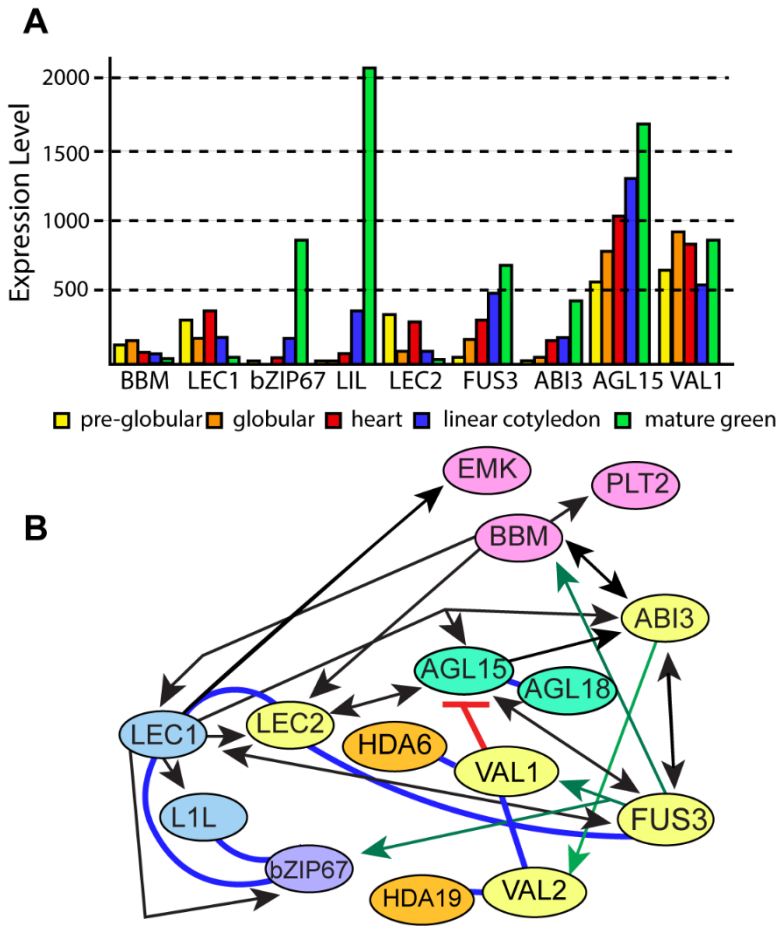



**Figure 1.2**. Interactions between key regulators of embryogenesis, the process of SE and the transition to seedling development. A: Transcript accumulation in embryos. Data is from the *Arabidopsis* eFP browser (Winter, Vinegar et al. 2007) from data provided by Goldberg and Harada. B: Working model summarizing interaction between select key regulators. Solid lines with arrows indicate a positive regulator loop. Other select direct targets are shown by black or green arrows. Direct responsive targets are from: LEC1 (Pelletier, Kwong et al. 2017); LEC2 (Braybrook, Stone et al. 2006); FUS3 (Yamamoto, Kagaya et al. 2010, Wang and Perry 2013); ABI3 (Mönke, Seifert et al. 2012, Tian, Wang et al. 2020); AGL15 (Zheng, Ren et al. 2009); BBM (Horstman, Li et al. 2017); VAL1/HSI2

(Chen, Veerappan et al. 2018). Protein interactions are diagrammed by blue connecting lines and are from: LEC1-LEC2 (Boulard, Thevenin et al. 2018); LEC2-FUS3 (Tang, Zhou et al. 2017); VAL2-HDA19 (Zhou, Tan et al. 2013); VAL1-HDA6 (Chhun, Chong et al. 2016); AGL15-AGL18 (Serivichyaswat, Ryu et al. 2015); LEC1/L1L-bZIP67 (Yamamoto, Kagaya et al. 2009). In addition, VAL1 and VAL2 can interact ((Chhun, Chong et al. 2016)(Chen, Wang et al. 2020)). Yellow, B3 domain; purple, basic leucine zipper; pink, AP2 domain; green, MADS domain; blue, NF-YB/HAP3; orange, histone deacetylase. Not all interesting interactions or proteins are shown, or relevant references cited in the interest of clarity.

## Chapter 2 Direct and indirect targets of the Arabidopsis seed transcription factor ABSCISIC ACID INSENSITIVE3

(This chapter has been published in The Plant Journal, 2020. I contributed to the analysis of the high-throughput data, and the follow-up experiments to verify and characterize targets.)

### 2.1 Introduction

Humans rely on seeds for the majority of our daily calories. Seed plants rely on seeds to perpetuate the species. Orthodox seeds are amazing structures that can allow the next generation of plants to survive in a dry state for long periods, even under adverse conditions (Roberts 1973). They perceive the environment and complete germination when survival to next seed set is optimal. Structures within the seed support the young plant until it becomes autotrophic. Angiosperm seed development begins with a double fertilization event, producing a diploid embryo and triploid endosperm. Maternal tissues of the ovule contribute the seed coat. Embryogenesis encompasses the processes from the one-celled zygote to the mature embryo and is often considered to occur in overlapping stages of morphogenesis and maturation (Braybrook and Harada 2008). During maturation, accumulation of storage products and preparation for desiccation occur, contributing to seeds' amazing properties. Other programs such as acquisition of dormancy may also occur. A network of transcription factors controls seed development with LEAFY COTYLEDON1 (LEC1), ABSCISIC ACID INSENSITIVE3 (ABI3), FUSCA3 (FUS3), and LEAFY COTYLEDON2 (LEC2) forming the LAFL network (reviewed in Boulard, Fatihi et al. (2017)). LEC2, FUS3 and ABI3 are all members of the B3 domain transcription factor family. LEC1 is a subunit of CCAAT-binding factors. Loss-of-function of any LAFL causes

loss of seed programs, especially programs involved in maturation, desiccation tolerance and dormancy. Some of the mutants (*lec1, lec2, fus3*) show a loss of embryo identity in that rescued homozygous embryos show partial conversion of cotyledons to leaves (reviewed in (Harada 2001)). Loss-of-function of *ABI3* does not show this phenotype, but it, like *lec1* and *fus3*, does not acquire desiccation tolerance properly and shows defects in storage product accumulation (Nambara, Naito et al. (1992). All can promote embryo programs to some extent when ectopically expressed via a *35S* promoter, with LEC1 and LEC2 showing ectopic embryos on the "seedling" upon completion of germination, and FUS3 and ABI3 showing cotyledon programs active in leaves (Parcy, Valon et al. 1994, Parcy and Giraudat 1997, Lotan, Ohto et al. 1998, Stone, Kwong et al. 2001). ABI3 is also involved in ABA perception and response (Nambara, Naito et al. 1992). The LAFL factors have been shown to interact genetically (To, Valon et al. 2006) and in some cases directly regulate each other (Braybrook, Stone et al. 2006, Wang and Perry 2013, Pelletier, Kwong et al. 2017).

The LAFL genes and another family of B3 domain proteins, namely VAL (VP1/ABI3-LIKE/HIGH LEVEL EXPRESSION OF SUGAR INDUCIBLE GENE 2 (VAL1/HSI2), VAL2/HSI2-LIKE1 (HSL) and VAL3), are involved in mediating the embryo to seedling transition (Suzuki, Wang et al. 2007, Suzuki and McCarty 2008, Veerappan, Wang et al. 2012). The *val1 val2* double mutant produces somatic embryos on seedlings and shows ectopic induction of the LAFL genes, as well as of *AGL15*, that encodes a MADS-domain transcription factor that promotes somatic embryogenesis (Harding et al., 2003; Thakare et al., 2008). Ectopic induction of *ABI3* was found to be involved in the development of

somatic embryos on seedlings in the *val1 val2* double mutant, while LEC2 and FUS3 were essential for this process (Jia, McCarty et al. 2013).

Prior work identified direct responsive targets of FUS3 and LEC1 using a combination of chromatin immunoprecipitation (ChIP) and tiling arrays (ChIP-chip), ChIP-seq and expression microarrays (Yamamoto, Kagaya et al. 2010, Wang and Perry 2013, Pelletier, Kwong et al. 2017). Putative direct targets for LEC2 were also determined, however the strategy involved induction of a LEC2-GR fusion protein to the nucleus by treatment with dexamethasone and assessment of transcriptomes at early time points, and did not involve ChIP (Braybrook, Stone et al. 2006). Mönke et al. (2012) published an ABI3 regulon derived from ChIP-chip and transcriptome experiments, however only 5% of the genes were identified as ABI3 targets in all three experiments (two ChIP arrays and one transcriptome study) and the minority of genes were identified as targets by at least two experiments. Less than ten percent of the regulon was identified being directly bound by ABI3 in the two ChIP-chip experiments, and several previously identified direct targets were not found as bound in their experiments, possibly due to challenges in acquiring robust ChIP from seeds (Haque, Han et al. 2018). Here we report on global targets for ABI3, using ChIP, an embryonic culture tissue, and the same tiling array as used for previous work with FUS3 (Wang and Perry 2013), allowing comparison between binding sites. We combine these data with transcriptome analysis to find direct responsive targets to ABI3 accumulation. Targets that respond to ABI3 accumulation but are not bound by ABI3 are putative indirect targets. Extensive interactions exist between key transcription factors in control of seed programs.

## 2.2 Results

### 2.2.1 Gene Expression in Response to ABI3

We used the GeneChip™ Arabidopsis 1.0 ST array (ThermoFisher) to assess the transcriptome in response to ABI3 accumulation. We compared developing seeds of *abi3-5* (Ooms et al., 1993) to Ler wild type (WT) at 15-16 days after flowering (daf). Using false discovery rate (FDR) cutoffs of <0.05 and at least a two-fold change, 884 genes showed increased transcript in 15-16 daf *abi3-5* seed compared to Ler WT and 627 genes showed reduced transcript accumulation for this comparison ((Tian, Wang et al. 2020) Data S1). PANTHER™ (Mi et al., 2016) was used to determine overrepresented gene ontology categories and select categories from "Biological Processes" are shown in Figure 2.1a and b respectively for ABI3 induced and repressed genes. Increased transcript accumulation from genes whose products are involved in plastid translation and chlorophyll biosynthetic processes in *abi3-5* (i.e., ABI3 repressed) are consistent with the embryos remaining green into maturation (Figure 2.1b) (Nambara, Naito et al. 1992, Delmas, Sankaranarayanan et al. 2013). Increased transcript accumulation from genes encoding proteins involved in starch catabolic processes (17.35 fold enrichment, FE, over the genome), glyoxylate cycle (20.49 FE), photosynthesis (6.53 FE), and other GO categories as shown in Figure 2.1b are consistent with the lack of establishment of a quiescent state in *abi3-5* (Nambara, Naito et al. 1992). Other overrepresented categories produce proteins associated with the lack of desiccation tolerance or dormancy in *abi3-5*. Genes encoding proteins involved in lipid storage were overrepresented in the list of genes with

reduced transcript accumulation in *abi3-5* consistent with a lack of accumulation of storage products in the mutant (Nambara, Naito et al. 1992).

### 2.2.2 Global Identification of In Vivo Binding Sites for ABI3

To map binding sites for ABI3, we used a ChIP-chip approach as previously performed for FUSCA3 (Wang and Perry 2013). *ABI3* was induced by its own promoter and included a 10x c-myc tag at the carboxyl terminal end to allow use of a commercial antibody for ChIP (Cell Signaling). The transgene was able to complement the *abi3-5* mutant phenotype (Figure 2.2). A *35Spro:AGL15* transgene in *abi3-5* was introduced as described in the Experimental Procedures, allowing us to initiate embryonic culture tissue (ECT) to easily obtain tissue for the ChIP studies as previously done for AGL15 and FUS3 (Wang, Tang et al. 2002, Harding, Tang et al. 2003). Although *35Spro:AGL15* was previously shown to induce endogenous *ABI3* transcript accumulation in whole developing seeds (Zheng et al., 2009), the ECT with the *ABI3pro:ABI3-myc/HA* transgene is not overaccumulating *ABI3* transcript compared to isolated developing zygotic embryos from the same line lacking the *35Spro:AGL15* transgene as shown and discussed in Figure 2.3.

Chromatin immunoprecipitation was performed as in Wang and Perry (Wang and Perry 2013). Probes were generated from three biological replicates to hybridize to Affymetrix GeneChip Arabidopsis Tiling 1.0 arrays. CisGenome (Ji, Jiang et al. 2008) was used to analyze the data with a moving average cutoff of 2.5 to match FUS3 analysis. This resulted in 2510 peaks assigned to 2278 genes ((Tian, Wang et al. 2020) Data S2; FDR < 0.03, using the left tail method in CisGenome; the closest gene was assigned to the peak).

From this point forward, ABI3 “associated genes” refers to these 2278 transcriptional units. The location of bound fragments relative to the nearest gene is summarized in Table 2.1.

Sequences from the bound peaks were analyzed using the Gibbs Motif Sampler (Ji, Jiang et al. 2008). Binding sites for bZIP and bHLH transcription factors called G-boxes (CACGTG; (Kawagoe and Murai 1996, Jakoby, Weisshaar et al. 2002)) were overrepresented. RY motifs (CATGCA; (Monke, Seifert et al. 2012)) that are binding sites for B3 domain proteins were not identified by Gibbs Motif Sampler as enriched in the 2510 regions bound by ABI3. However, when compared to three matched sets of randomly generated sequence with the same average peak size (i.e., 829 bp, generated in CisGenome), RY motifs, both long and short, were significantly enriched in the ABI3 bound regions, as were G boxes (Table 2.2, chi-square test $P<0.0001$).

To determine completeness of our ABI3 bound dataset, we looked at whether previously reported targets were included in our list. As shown in Table 2.3, ABI3 targets identified by different approaches were bound in our ChIP-chip experiment (peak number reflects the rank of the peak with peak 1 as the strongest peak). *PP2A (At1g69960), TUB2 (At5g62690)* and *EF1ALPHA (At5g60390)* used as non-bound controls in Lim et al. (Lim, Park et al. 2013), Xi et al. (Xi, Liu et al. 2010) and Bedi et al. (Bedi, Sengupta et al. 2016), respectively were not bound in our experiment. Thus, the ChIP-chip data we collected for ABI3 captured previously identified direct targets.

Mönke et al. (Monke, Seifert et al. 2012) reported on the ABI3 regulon. This set of genes was compiled from two different ChIP arrays as well as transcriptome studies and resulted in 320 genes putatively directly regulated by ABI3. Of our 2278 genes with regulatory regions associated with ABI3 in vivo, 60 were identified as direct targets using the SAP chip experiment (Monke, Seifert et al. 2012). This overlap includes our peak 2366 ((Tian, Wang et al. 2020) Data S2). Nineteen of these genes also were identified as direct ABI3 targets on the ChIP-chip experiment using the Agilent array (Monke, Seifert et al. 2012) and this included our peak 1985. This boosts our confidence that even the "weaker" peaks identified in our study are true binding sites as they were identified in three separate experiments. It also indicates overlap for ABI3 binding in embryonic culture tissue and seeds. ChIP studies in Mönke et al. (Monke, Seifert et al. 2012) were missing targets previously identified as direct in seeds, such as *SOM* and *MFT1*. This may be due to the challenges of acquiring robust ChIP populations from developing seeds, especially oil accumulating seeds (Haque, Han et al. 2018).When including the transcriptome study, 49% of the genes reported as in the ABI3 regulon by Mönke et al. (Monke, Seifert et al. 2012) are bound by ABI3 in our study ((Tian, Wang et al. 2020) Data S2).

Bargmann et al. (Bargmann, Marshall-Colon et al. 2013) developed TARGET (Transient Assay Reporting Genome-wide Effects of Transcription factors) to rapidly assess direct targets of ABI3. By inducing nuclear localization of ABI3-GR in protoplasts by addition of dexamethasone in the presence of cycloheximide to block translation, they identified 186 induced and 272 repressed targets that are potentially directly regulated by ABI3. Ninety-eight of the induced genes (53%) were bound by ABI3 in our ChIP-chip

study ((Tian, Wang et al. 2020) Data S2). However, only 18 (6%) of their directly repressed list were found to be bound in our study.

Recently, a publication reported on direct targets of *Glycine max* (Gm)ABI3. We compared our putative ABI3 bound sites to potential Arabidopsis orthologs in soybean and found that ~30% of genes bound by ABI3 in Arabidopsis are also bound by GmABI3 (Jo et al., 2020; (Tian, Wang et al. 2020) Data S2). This includes our peak 2507, again supporting our confidence that even weaker peaks are true binding sites relevant to zygotic embryos.

### 2.2.3 Direct Responsive Target Genes of ABI3

Data comparing the transcriptome of *abi3-5* to Ler developing seeds was combined with the ChIP-chip results to differentiate direct responsive from indirect responsive targets. With the cutoffs we used, we generated lists of 317 ABI3 directly induced target genes (i.e., reduced transcript accumulation in *abi3-5* compared to WT Ler and associated with ABI3) and 87 ABI3 directly repressed target genes ((Tian, Wang et al. 2020) Data S3). Select gene ontology categories significantly overrepresented for the ABI3 direct induced and repressed lists are shown in Figure 2.1a and 2.1b, respectively.

Interestingly, there were differences in *cis* motifs between direct induced and direct repressed genes (Table 2.2). Short- (CATGCA) and long-RY motifs (CATGCATG) that serve as binding sites for B3 domain proteins such as ABI3 were significantly overrepresented in bound regions for direct induced genes compared to the average of three matched sets with the same mean peak width (1368 bp). However, these motifs

were not overrepresented in the bound regions for the direct repressed genes. G-boxes that serve as DNA binding sites for basic helix loop helix containing proteins such as PIL5/PIF1, as well as bZIP transcription factors such as ABI5, that have been found to interact with ABI3 (Nakamura, Lynch et al. 2001, Park, Lee et al. 2011) were overrepresented in the total bound fragments. G-boxes were also significantly overrepresented in the DNA sequences bound in both the direct induced and the direct repressed datasets, with a statistically significant increase in the induced compared to repressed lists (Table 2.2). Another protein reported to interact with ABI3 in control of gene expression is the AP2 domain protein, ABI4, that binds CACCG (Nakamura, Lynch et al. 2001, Huang, Zhang et al. 2017). This motif was significantly overrepresented in the DNA sequences corresponding to direct induced genes, but not in direct repressed genes.

### 2.2.4 Overlap with Genes Regulated by FUS3 and LEC2

Because ABI3, LEC2 and FUS3 are all B3 domain DNA-binding proteins involved in seed development, we were interested in determining regions of the genome that are bound by more than one of these transcription factors. We compared the ChIP-chip data for FUS3 from Wang and Perry (Wang and Perry 2013) to the ABI3 data reported here. Both datasets were generated from ECT so differences in binding were not due to seed age. For FUS3, 1218 binding sites assigned to 1140 genes were reported, whereas for ABI3, this numbers 2510 sites (2278 genes) using the same cutoffs in CisGenome. A little more than three hundred (313) genes were bound by both ABI3 and FUS3 (354 binding sites for ABI3 and 338 for FUS3 – some genes are associated with two binding regions for one or the other factor). GO analysis showed these 313 genes were overrepresented for the

categories seed development, response to karrikin, transcription, and response to hormone (Figure 2.4a). We then determined whether the binding sites partially or completely overlapped for FUS3 and ABI3 or if these factors bound different regions of the gene. More than one third (132/354 sites or 37%) showed overlap such that the binding peak for one factor completely overlapped with the other factor. When also considering partial overlap of sites, this number rose to 221 or 62% (corresponding to 219 genes). Nearly all of these regions (207) had at least one CATGCA motif (94% of the 221 regions) and 20% of the associated genes were responsive to both ABI3 and FUS3 at FDR<0.05. The set of genes with overlapping FUS3 and ABI3 binding sites were overrepresented for seed oil body biogenesis, lipid storage, and seed dormancy (Figure 2.4a). Another set of genes were bound by FUS3 and ABI3 but the binding sites did not overlap. This constitutes a set of 133 ABI3 binding regions (122 genes; 38% of the total that may be coregulated). Only 35% of these regions had a CATGCA motif and only 17 (13%) of the associated genes responded to both FUS3 and ABI3 accumulation. Overrepresented GO categories for this set of genes included response to ethylene and response to water deprivation. Twenty-eight genes had associated binding profiles such that there were shared and distinct binding sites for FUS3 and ABI3. No significant GO enrichment was found among these genes.

Forty-three genes were not only bound by both FUS3 and ABI3 but were also induced in response to these transcription factors. This set showed overrepresentation of genes involved in lipid storage (Figure 2.4b). Genes that were directly induced by ABI3, but were not direct targets of FUS3 (Wang and Perry 2013) numbered 273 genes and were

overrepresented for response to ABA, seed development, response to water deprivation, and other processes (Figure 2.4b). Genes directly bound by, and responsive to, FUS3, but not bound by ABI3 numbered 177 and showed overrepresentation for somatic embryogenesis, lipid storage, regulation of seed maturation, and other processes (Figure 2.4b). Although genes in seed development were overrepresented in both ABI3 and FUS3 unique direct targets, they were different genes. There are an additional seven genes involved in seed development that were direct induced targets of both B3 domain proteins, but the Fold Enrichment (FE) of 6.55 had an FDR=0.08.

When considering genes directly bound by, and responsive to, FUS3 and ABI3 but broadening the number of genes by using $p<0.05$ and 1.5 fold cutoffs for ABI3, 78 genes were regulated by both. Thirteen of these genes may also be directly induced by LEC2 based on the observation that they showed increased transcript accumulation within 4 hours of dexamethasone treatment of seedlings with a LEC2-GR fusion protein (Braybrook et al., 2006). Five of the thirteen genes in this category encode transcriptional regulators (Table 2.4). Two additional gene products were predicted to localize to the nucleus. The majority of the remainder were involved in storage product accumulation (Table 2.4).

### 2.2.5 ABI3, FUS 3 and the Regulation of *Pri-miR156* Encoding Genes

Prior work documented that FUS3 was able to directly induce genes involved in production of *MIR156* (Wang and Perry 2013). *MIR156* is involved in preventing precocious maturation (Rodrigues and Miguel 2017) and, after completion of germination, other developmental phase changes that are related to FUS3 (Lumba, Tsuchiya et al. 2012). We proposed that FUS3, that appears to only induce direct target genes, is involved

in indirect repression of the *MIR156* targets *SPL10* and *SPL11* via MIR156 to prevent premature activation of maturation genes during seed development as well as documented phase transitions that occur in the *fus3* loss-of-function (Nodine and Bartel 2010, Willmann, Mehalick et al. 2011, Lumba, Tsuchiya et al. 2012). Here we find that ABI3 also directly regulates expression of genes that produce *MIR156*. Data from ChIP-chip indicates direct binding of ABI3 to regulatory regions of the *pri-miR156A, C* and *D* genes, but not the MIR156 targets *SPL10* and *SPL11* (Figure 2.5). We verified association of ABI3 with *pri-miRNA D* that was the "weakest" target of the three using ABI3 tagged with a different epitope than used for ChIP (HA tag, Figure 2.6a and ChIP with myc tag Figure 2.5). In addition, we found that during an early stage of seed development (7-8 dap), transcript accumulation was decreased in *abi3-5* compared to Ler WT, indicating ABI3 induces expression of these genes (Figure 2.6b). Transcript accumulation at 11-12 dap was also reduced for *pri-miRNA156A* and *C*, but *pri-miRNA156D* no longer had a significant change. During later seed development (13-14 and 15-16 dap), ABI3 represses transcript accumulation from *pri-miRNA156D* as documented by an increase in these transcripts in *abi3-5* compared to Ler WT (Figure 2.6b). Consistent with this result, a decrease in transcript accumulation of *SPL10* and *SPL11* was observed at 15-16 dap in *abi3-5* compared to Ler WT (Figure 2.6c). As previously reported FUS3 induces *pri-miRNA156A* and *C* at 11-12 dap (Wang and Perry, 2013; Figure 2.6d), as well as at a later stage of seed development (13-14), but there were no significant changes in transcript accumulation of *pri-miRNA156D* at either stage. As shown in Figure 2.7, if one examines a timecourse of accumulation of *pri-miRNA156* within a genotype, the transcript

accumulation of all three microRNA encoding genes does not show any significant changes in Ler, WT, but showed significant increase as the seeds developed in *abi3-5*.

### 2.2.6 ABI3, Seed Desiccation Tolerance and Longevity

One phenotype of *abi3* is the lack of acquisition of desiccation tolerance. González-Morales et al. (Gonzalez-Morales, Chavez-Montes et al. 2016) performed an analysis of desiccation tolerant and intolerant Arabidopsis lines to generate a regulatory network during acquisition of seed desiccation tolerance. This regulatory network included transcription factors (TFs) that formed two major co-expression subnetworks that they named TFsSeed-subNetDT1 and TFsSeed-subNetDT2. While it is unsurprising that the genes encoding these TFs are responsive to ABI3 accumulation since *abi3-5* was one of the mutants used in the study, what is surprising is how many of the TFs are directly regulated by ABI3. All eleven of the TFs in TFsSeed-sNetDT2 and six of the nine TFs in TFsSeed-sNetDT1 were direct targets of ABI3. The majority showed significant activation of expression in response to ABI3 accumulation at the cutoffs we used. These TFs included PLATZ1, PLATZ2 and AGAMOUS LIKE-67 (AGL67) that were shown to have key roles in desiccation tolerance (Gonzalez-Morales, Chavez-Montes et al. 2016). Two of the three genes in TFsSeed-sNetDT1 that were not direct targets of ABI3, are direct targets of FUS3, with a third one appearing to be an indirect target (Wang and Perry 2013). Genes coexpressed with these TFs led to development of FullSeed-subNetDT1 and subNetDT2, with DT1 mainly involved with seed filling early (15 daf), and with desiccation tolerance, along with DT2 at later stages (17 and 21 daf). More than one half of the genes in these FullSeed subnets are directly bound by ABI3 (57% for DT1 and 53% for DT2; (Tian, Wang

et al. 2020) Data S4). Most of these are expressed in response to ABI3 accumulation (111 of the genes in DT1 and 107 of the genes in DT2). Only three are repressed.

The Late Embryogenic Abundant (LEA) proteins are associated with organisms and tissues that undergo dehydration and are thought to be involved in protection of cellular components during desiccation (reviewed in (Hundertmark and Hincha 2008)). Thirty of the fifty-one LEA proteins in Arabidopsis had regulatory regions associated with ABI3 and twenty-two showed significant transcript reductions in *abi3-5* compared to Ler WT (Figure 2.8). Thus, over 40% of genes encoding LEAs were direct induced ABI3 targets. All of the genes in the LEA_4 group that are induced in seeds were direct targets of ABI3 and all but one were induced in response to ABI3. Those not directly associated are induced in non-seed tissues (Figure 2.8, (Hundertmark and Hincha 2008)). Likewise, all of the genes encoding dehydrins and LEA_5 genes that are induced in seeds were directly up-regulated in response to ABI3 accumulation with the exception of *At3g50970*. In contrast, most genes in the LEA_2, LEA_3 and SMP groups were not directly regulated by ABI3.

Desiccation tolerance is linked to seed longevity, and genes mentioned above are involved in both processes. For example, *LEA14* (*At1g01470*)*, XERO1* (*At3g50980*), *RAB18* (*At5g66400*), *AtEM1 (At3g51810)* and *ATEM6 (At2g40170)* (EM:early methionine labeled) are all ABI3 directly induced LEAs and these genes have been implicated in seed longevity (reviewed in (Sano, Rajjou et al. 2016) (Carbonero, Iglesias-Fernandez et al. 2017)).

Raffinose family oligosaccharides (RFOs) are thought to be involved in longevity (reviewed in (Leprince, Pellizzaro et al. 2017)) and genes encoding GALACTINOL

SYNTHASE2 (*At1g56600*) and STACHYOSE SYNTHASE (*At4g01970*) were directly bound by ABI3 and induced in response to ABI3 accumulation. *SEED IMBIBITION PROTEIN2*, that was originally predicted to be a raffinose synthase, was found to be a raffinose-specific α-galactosidase involved in breakdown of raffinose to sucrose and galactose (Peters, Egert et al. 2010). This gene is associated with an ABI3 binding site ~2.4 kb upstream of the TSS (assigned to *At3g57530*) and an increase in transcript 5.3 fold in *abi3-5* seeds compared to Ler WT

Lack of chlorophyll degradation is thought to be related to decreased longevity (reviewed in (Leprince, Pellizzaro et al. 2017)) and one phenotype of *abi3* is that the seeds remain green at maturity, in part due to the lack of expression of *NYC1* (*NON-YELLOW COLORING1*) and *NOL* (*NON-YELLOW COLORING-LIKE*) that are involved in chlorophyll breakdown. *NYC1* was a direct induced ABI3 target. STAY-GREEN2 (*SGR2*), is also involved in chlorophyll breakdown and was also a direct ABI3 induced target (Leprince, Pellizzaro et al. 2017). Other direct ABI3 induced targets reported to be involved in seed longevity include HEAT STRESS TRANSCRIPTION FACTOR A9 (HSFA9; (Kotak, Vierling et al. 2007)) and the aquaporin TONOPLAST INTRINSIC PROTEIN3;1 (TIP3;1, reviewed in (Carbonero, Iglesias-Fernandez et al. 2017)).

During aging, oxidation of macromolecules occurs that can lead to loss of viability (reviewed in (Sano, Rajjou et al. 2016)). A number of mechanisms help protect against this damage including seed storage proteins (Nguyen, Cueff et al. 2015) and antioxidants (Sattler, Gilliland et al. 2004). Two cruciferin genes and two seed storage albumin genes were directly induced ABI3 genes. The VITAMIN E locus 1 (*VTE1*), which is involved in the

limiting step of tocopherol biosynthesis and *TEMPERATURE INDUCED LIPOCALIN* (*AtTIL*), producing lipocalin were also directly induced genes.

Another aspect to viability is the ability to repair damage accrued during aging (reviewed in (Sano, Rajjou et al. 2016)). While genes encoding products involved in nucleic acid repair were not represented in the direct ABI3 genes, those involved in protein repair were directly induced by ABI3 including two genes encoding METHIONINE SULFOXIDE REDUCTASES (MSRs, *At4g04840* and *At1g53670*), and the second L-ISOAPARTYL O-METHYLTRANSFERASE (*PIMT2*, *At5g50240*; Figure 2.9).

### 2.2.7 ABI3 and Seed Dormancy

Carrera et al. (Carrera, Holman et al. 2007) compared Ler dormant to afterripened seeds and their dataset overlaps extensively with ABI3 direct targets (Table 2.5), indicating that ABI3 directly regulates many genes associated with dormancy. Two loci were first identified as QTLs affecting dormancy, *DELAY OF GERMINATION1 (DOG1, At5g45830*) and *REDUCED DORMANCY5* (*RDO5, At4g11040*) (reviewed in (Chahtane, Kim et al. 2017)). Both of these genes are directly induced targets of ABI3 (peak 341 and 196 respectively) and FUS3 (peak 567 and 517 respectively; (Wang and Perry 2013)). While the transcript accumulation of *RDO5* did not meet the cutoff of 2-fold, (*abi3-5/Ler* = 0.75) the decreased transcript accumulation in *abi3-5* was significant at FDR<0.05.

Liu et al. (Liu, Zhang et al. 2013) demonstrated that auxin promotes seed dormancy through AUXIN RESPONSE FACTORS (ARFs) mediated induction of *ABI3*. Among other components, *ARF10* (*At2g28350*) and *ARF16* (*At4g30080*) show decreased

dormancy in loss-of-function mutants. These genes are posttranscriptionally regulated by *MIR160*. ARF10 and ARF16 upregulate *ABI3*, likely indirectly (Liu et al., 2013). The microarray data in the study presented here showed that both *ARF10* and *ARF16* transcript accumulation were significantly decreased in *abi3-5* (0.52 and 0.60 compared to Ler WT, respectively). The gene encoding *MIR160B* (*At4g17788*) showed seventeen fold increased transcript accumulation in *abi3-5* and appeared to be a directly regulated target of ABI3 (peak 424 was assigned to *At4g17790* but is actually between this gene and the *MIR160B* encoding gene, Figure 2.10a). *ARF10* and *ARF16* could be direct ABI3 targets as weak binding sites are present at the 5' end of these genes but they did not meet our cutoffs in CisGenome (Figure 2.10b and 2.10c). We confirmed that ABI3 associates with regulatory regions of the gene encoding *MIR160B* by ChIP-qPCR (Figure 2.11a), using ECT accumulating ABI3-HA. The average from the three biological replicates was not significantly different from the non-bound control due to the wide range of fold change in the individual experiments (2.1 to 17.2) but was significantly different within each experiment, indicating association of ABI3 with *At4g17788* regulatory regions. We also performed ChIP with developing siliques expressing *ABI3-myc* from the native promoter and found a fold change greater than two in two biological replicates. In addition, three independent biological replicates assessing transcript in 15-16 daf seeds showed a significant increase in transcript from this gene in *abi3-5* compared to Ler WT (Figure 2.11b). Meanwhile, *ARF10* and *ARF16* had significantly decreased transcript in *abi3-5* compared to Ler WT (Figure 2.11b). Thus, there appears to be a feedback loop whereby ABI3 directly represses the gene encoding *MIR160B*, allowing an increase in transcript of

*ARF10* and *ARF16* that can in turn up regulate ABI3, with a net effect resulting in seed dormancy (Figure 2.9).

### 2.2.8 Roles of ABI3 Outside of Seed Development

Costa et al. (Costa, Righetti et al. 2015) performed an analysis looking at the re-establishment of desiccation tolerance in Arabidopsis seedlings that had protruded the radicle and therefore completed germination. They found that two hours of ABA incubation was insufficient to reestablish desiccation tolerance in these germinated seeds, but by 24 hours of ABA treatment, there was 100% desiccation tolerance. Among differentially induced genes were *At4g25580*, *COLD ACCLIMATION PROTEIN160* (*CAP160*), a LEA group 4 protein *EMBRYONIC CELL PROTEIN63* (*ECP63*; *At2g36640*) and *CYSTATHIONINE BETA-SYNTHASE4* (*CBSX4*; *At1g80090)*. All three of these genes were directly bound, included targets of ABI3 (Monke, Seifert et al. 2012) and (Tian, Wang et al. 2020) Data S3). However, Costa et al. (Costa, Righetti et al. 2015) found that loss-of-function of these genes led to improved desiccation tolerance that appears contradictory to regulation by ABI3. They proposed redundant and compensatory gene regulation. In fact, the closest homologs of *At4g25580, At5g52300* and *ECP63, At3g53040* were also directly induced targets of ABI3 in our dataset.

To look more broadly at potential overlap of the ABI3 regulon during seed development and re-establishment of desiccation tolerance, the data from Costa et al. (Costa, Righetti et al. 2015) (GSE62876) was analyzed in GEO2R using default settings (Benjamini and Hochberg –FDR) and lists with cutoffs of FDR<0.05 and at least a two-fold change were generated for 0 hours ABA treatment compared to 2, 12, and 24 hours ABA

treatment. As shown in Table 2.6 and (Tian, Wang et al. 2020) Data S5), there was a high degree of overlap between genes directly induced by ABI3 in our study and those induced in response to ABA during redesiccation. For the ABI3 direct induced genes (317), 56, 59 and 62% were also upregulated in response to ABA treatment at 2, 12, and 24 hours respectively. Only 4-6% of the ABI3 direct induced genes were repressed by ABA treatment to reestablish desiccation tolerance. For ABI3 indirect induced genes, a lower fraction overlapped with the ABA induced genes. ABI3 shows increased transcript accumulation during ABA treatment with a 2.8-fold increase at 2 hour ABA compared to 0 hour ABA. Therefore, ABI3 may be involved in directly regulating many of the genes involved in reestablishment of desiccation tolerance after the completion of germination. ABI3 directly and indirectly repressed genes do not show the degree of overlap as ABI3 directly induced genes (Table 2.6). Very few genes, ABI3 direct or indirect, induced or repressed, overlapped with genes differentially induced between 24 and 72 hours ABA treatment, a period of time that had an impact on longevity, as measured by accelerated aging (Costa, Righetti et al. 2015).

ABI3 has been reported to be involved in response to dehydration stress in 3-4 week old seedlings, with *ABI3* transcript and protein accumulation increasing during dehydration and remaining at higher levels during recovery (Bedi, Sengupta et al. 2016). These authors demonstrated direct regulation by ABI3 of seven genes, all of which were present in the ChIP-chip data using ECT. To look more broadly at potential regulation during drought stress by ABI3, we compared ABI3 direct targets to genes showing differential transcript accumulation during moderate and progressive drought stress in

25-35 day seedlings (Harb, Krishnan et al. 2010); (Tian, Wang et al. 2020) Data S5). Of the 317 ABI3 direct induced targets in developing seeds, 155 (49%) were responsive to drought in seedlings with 130 (41%) showing increased transcript accumulation at some stage of drought stress. Of the 87 ABI3 direct repressed targets in developing seeds, 59 (68%) responded to drought in seedlings with about equal numbers showing increased transcript (27 or 31%) compared to decreased transcript (32 or 37%). Thus, potential exists for more extensive regulation by ABI3 during drought stress.

### 2.2.9 Potential for Cross-regulation Among ABA Regulated Transcription Factors

Song et al. (Song, Huang et al. 2016) performed an extensive ChIP-seq study of genome wide binding by twenty-one ABA responsive TFs, and RNA-seq study of ABA response in 3 day etiolated seedlings to elucidate the ABA regulatory network. Although ABI3 was not included in their study, nine of the twenty-one ABA associated TFs that were included were direct regulated targets of ABI3 in our study. These include the products of *ANAC032, ANAC102*, *DREB2A, GBF2* and *GBF3* that were ABI3 directly upregulated targets. Some did not make our cutoffs for direct induced targets because the ratio of transcript in *abi3-5* compared to Ler developing seeds did not fall below 0.5, but changes in transcript were significant. Another four TFs in their study were directly repressed targets of ABI3 (*RD26, FBH3, HB5, HB7*; (Song, Huang et al. 2016). Interestingly, of these, FBH3 and HB5 were the only TFs among the twenty-one that do not show increased number of binding sites after ABA treatment. Of the nine directly bound and responsive to ABI3, only these two and DREB2A do not bind regulatory regions associated with ABI3. Regulatory regions for eight additional TFs were associated with ABI3, but did not appear to be

induced in developing seeds. Although very young (3d) etiolated seedlings were used (Song, Huang et al. 2016), binding sites likely overlap with developing seeds. To extend the ABI3 network from the ABI3 directly induced ABA TFs to ABI3 indirectly regulated targets, we retrieved target genes for the five ABI3 direct induced TFs from www.ABAtf.net. (Song, Huang et al. 2016) and compared these lists to the ABI3 direct/indirect and induced/repressed gene sets. More than one-half of the ABI3 indirectly induced (241 of 311, 77%) and ABI3 indirectly repressed (546 of 797, 69%) genes were directly bound by at least one of the five ABI3 direct induced TFs (Tian, Wang et al. 2020); Data S6), suggesting that the majority of ABI3 indirect targets may be regulated by these five ABI3 direct induced TFs. An even larger fraction of ABI3 direct targets (90% of the 317 directly induced and 95% of the 87 directly repressed) were also bound by at least one of these five TFs (Tian, Wang et al. 2020) Data S6), indicating the potential for extensive interactions with ABI3 in gene regulation.

## 2.3 Discussion

### 2.3.1 Numbers and Types of Genes Regulated by ABI3

We determined direct and indirect targets of the TF, ABI3. Based on mutant phenotypes, ABI3 is involved in seed maturation, establishment of a quiescent state and degreening of the embryo (Nambara, Naito et al. 1992, Delmas, Sankaranarayanan et al. 2013). The types of genes controlled by this TF reflect these processes. Only ~18% of the ABI3 direct targets responded to ABI3 accumulation in developing seeds at the cutoffs we used. This is comparable to the fraction of FUS3 associated genes that showed response to FUS3 (19%; Yamamoto et al., 2010; (Yamamoto, Kagaya et al. 2010, Wang and Perry

2013). Although this represents a minority of sites, many genes were significantly responsive to ABI3 accumulation at lower cutoffs and 18% is within the range of what has been found in other studies (e.g. (Lee, He et al. 2007, Oh, Kang et al. 2009).

While the Gibbs motif sampler identified RY motifs and G-boxes as overrepresented among the sites bound by FUS3 (Wang and Perry 2013), only the G-box was identified by this method for sites bound by ABI3. ABI3 has been shown by protein binding arrays to bind to a somewhat less conserved RY motif than FUS3 with only the core CATG absolutely required (Baud et. al., 2016). A more recent study using EMSA extended the motif to CATGCA with more flexibility at the 3' position (Sasnauskas, Manakova et al. 2018). Greater flexibility outside of the core sequence may be why the Gibbs motif sampler did not identify RY motifs among the in vivo binding sites for ABI3. However, when comparing matched sets of DNA sequences, RY motifs were overrepresented with 40% of the binding regions containing at least one RY motif of form CATGCA and 6% of form CATGCATG. These were significant increases over the average of the matched sets that had 26% CATGCA and 2% CATGCATG, but lower than the fraction of sites with these motifs that FUS3 bound (75% and 22% respectively, Wang and Perry, 2013). Conversely, G-boxes were increasingly found in ABI3 binding sites compared to FUS3 sites with 51% and 12%, respectively of sites bound having at least one CACGTG site (Table 2.4 and (Wang and Perry 2013). This may reflect proteins with which these factors interact and possibly the fact that this motif contains the core ABRE of ACGT (Nakashima and Yamaguchi-Shinozaki 2013). While both ABI3 and FUS3 have a B3 domain that is involved in binding RY motifs, ABI3 has a B1 and B2 domain whereas FUS3 only has a B2

domain. These domains are involved in interaction with bZIP proteins and ABI3's B1 domain is necessary for interaction with ABI5 (Nakamura, Lynch et al. 2001). Several bZIP and bHLH TFs that bind to G-boxes or variants, have been identified that interact with ABI3, but have not been found to interact with FUS3 (Biogrid, https://thebiogrid.org/; (Chatr-Aryamontri A, Oughtred et al. 2017). While ABI3 has been shown to bind DNA (Park, Lee et al. 2011), it is possible that a significant fraction of the interactions are via an associated bZIP or bHLH and a G-box. The ChIP-chip study presented here would not differentiate among these possibilities.

### 2.3.2 ABI3 can both Activate and Repress Direct Target Genes

ABI3 appears to function mainly as a transcriptional activator, although it also directly repressed some genes. In contrast, FUS3 only directly activates gene expression (Wang and Perry 2013). Both genes have predicted transactivation domains (Monke, Altschmied et al. 2004, Kagaya, Okuda et al. 2005). ABI3 contains a putative ERF-associated amphiphilic repression (EAR) domain of LxLxL form that is conserved in the maize ortholog VP1 with a sequence of LKLIL in both species. This type of motif has been shown to recruit histone deacetylase complexes to repress gene expression (Ohta, Matsui et al. 2001) and may explain ABI3's ability to directly repress gene expression. FUS3 does not contain this motif.

There are also some intriguing differences in *cis* motifs overrepresented in the ABI3 bound regions of direct induced compared to direct repressed genes. The *cis* motif that B3 domain proteins recognize was only overrepresented in the DNA sequences controlling the direct induced genes. Likewise, *cis* elements for binding of ABI3 interacting

protein ABI4 were also only overrepresented in the direct induced gene set. G-box binding sites for bHLH/bZIPs were present in both induced and repressed regulatory regions. This may indicate different modes of binding for ABI3 to activate or to repress a gene. Possibly a direct ABI3-DNA interaction via a RY motif, with possibly interaction also of ABI4 or other AP2 domain protein with CACCG leads to gene activation. More than one-third (44%) of the DNA sequences bound for the direct induced genes have at least one RY motif and at least one CACCG motif compared to only 8% of the direct repressed regulatory sequences. G-boxes could be involved in repression, but they are also overrepresented in the induced dataset.

### 2.3.3 Overlap of Gene Regulation between LAFL and VAL Transcription Factors

Relatively few genes appear to be regulated by all three embryo B3 domain TFs (Table 2.4), but more than half of those that are function as TFs or are otherwise localized to the nucleus. LEC2 target identification involved induction of nuclear localization of LEC2-GR and measurement of transcriptomes at short (1 and 4 hour) time-points (Braybrook, Stone et al. 2006) and did not reveal the location of binding of LEC2. ABI3 and FUS3 involved ChIP-chip using the same platform and similar tissue (ECT; the difference being the epitope tagged factor expressed and the background ecotype). Combined with microarray data, this allows determination of overlap of gene regulation as well as potential overlap of binding sites. One-third of the genes potentially regulated by both FUS3 and ABI3 did not have overlapping binding sites for these proteins and there were some differences in GO categories for genes overrepresented among the different binding profiles (Figure 2.6). Some information may be lost when incorporating the

expression data because the stages of development were different in the *fus3*-3 (Yamamoto, Kagaya et al. 2010) and *abi3-5* studies. However, a larger percentage (87%) of shared direct responsive targets had at least partial overlap of binding sites between FUS3 and ABI3. Very few genes directly or indirectly regulated by ABI3 showed contradictory pattern of transcript accumulation in response to FUS3 or ABI3 (Tian, Wang et al. 2020) Data S7). Of the 317 ABI3 direct induced genes, 194 (61%) had a fold change comparing *fus3-3* to Col WT 12 daf developing seed of 0.5 or less, indicating induction of expression, direct or indirect, in response to FUS3 accumulation. For ABI3 indirect induced genes, 143 (46% of ABI3s 311 indirect expressed genes) were also induced in response to FUS3. For the 87 ABI3 direct repressed genes, two (2%) were induced and four (5%) were repressed in response to FUS3 accumulation (direct and indirect FUS3 targets). For ABI3 indirect repressed genes, these numbers were seven (2%) and thirty-six (12%) respectively.

The ChIP-chip and microarray results indicated that ABI3 directly induced over 40% of the LEA genes with all or nearly all seed expressed LEA's in groups 4, 5 and the dehydrin group included as direct targets. FUS3 accumulation also influenced many of the LEA genes with twenty-five responsive at 12 daf (Yamamoto et al., 2010). However, only 2 LEA genes showed direct association with FUS3 (Wang and Perry 2013). Because the tissue used for FUS3 and ABI3 ChIP-chip was ECT rather than staged seeds, the difference in binding cannot be attributed to different stages of development. Thus, LEAs appear to be largely under the control of ABI3 rather than FUS3.

Like FUS3 (Wang and Perry 2013), ABI3 directly induced *WRI*, which encodes a TF involved in regulating genes involved in fatty acid synthesis. Also, as previously reported for FUS3, ABI3 also directly regulated genes encoding products that generate fatty acids, but with the exception of *KCS18/FAE1* (*At4g34520*), the other genes do not overlap with direct FUS3 targets. While FUS3 directly regulates *ACP1 (AT3g05020)* and *LACS4 (AT4g23850)*, ABI3 directly induced *FAD3 (At2g29980)* and *SSI2 (At2g43710)*.

Another group of B3 domain TFs, VAL1/HSI2, VAL2/HSL1 and VAL3 function to promote the transition from embryo to seedling development by shutting down the AFL network (Suzuki, Wang et al. 2007, Suzuki and McCarty 2008, Veerappan, Wang et al. 2012), possibly by direct repression of AGL15 by VAL1/HSI2. (Zheng, Ren et al. 2009, Chen, Veerappan et al. 2018). However, while a *35Spro:AGL15* transgene promotes embryo tissue development from cultured embryos and seeds that complete germination in the presence of 2,4-D (Harding, Tang et al. 2003, Thakare, Tang et al. 2008), relatively normal seedlings develop without 2,4-D, indicating other components are involved. Interestingly, *VAL1/HSI2* was found to be a direct induced target of FUS3 (Wang and Perry 2013), while *VAL2/HSL1* appeared to be directly targeted and induced by ABI3 (the peak for *VAL2* was assigned to the neighboring gene). ABI3 and FUS3 autoregulate themselves as well and directly regulate each other. FUS3 directly induces *BBM* (Wang and Perry, 2013), and in turn BBM directly induces *LEC1*, *LEC2* and *ABI3* (Horstman, Li et al. 2017). LEC1 and LEC2 can form a complex to control gene expression, and it is possible that LEC1 can also interact with ABI3 (Boulard, Thevenin et al. 2018). Interactions of select embryo TFs are diagrammed in Figure 2.12.

We previously found much overlap between genes derepressed in *val1 val2* double mutant seedlings (88) and genes directly induced by FUS3 (220 - 40%; (Wang and Perry 2013). Very few genes that had decreased transcript in *val1 val2* (therefore VAL1 VAL2 induced) were directly induced FUS3 targets (i.e., five). A similar case is true for ABI3 with 32% overlap between ABI3 direct induced and *val1 val2* derepressed genes (i.e., increased in *val1 val2* compared to WT, indicating that VAL1 VAL2 repress these genes in seedlings), but only 0.9% overlap between ABI3 direct expressed induced and VAL1 VAL2 expressed genes (Tian, Wang et al. 2020) Data S7). Thus, like FUS3 and VAL1 VAL2, ABI3 and the VALs act with opposite consequences on many genes in embryos and seedlings. Transcriptomes in response to the single *val1* loss-of-function in developing embryos was also previously assessed (Schneider, Aghamirzaie et al. 2016). In this case, VAL1 appears to act coordinately as much as antagonistically with ABI3 with 21% and 14% of the ABI3 direct expressed genes induced and repressed, respectively by VAL1. For ABI3 directly repressed genes, 23% are repressed in embryos in response to VAL1 whereas 11% are induced (Tian, Wang et al. 2020) Data S7).

### 2.3.4 Transition from Morphogenesis to Maturation

Both FUS3 and ABI3 regulate members of the gene family that encodes MIR156. This microRNA regulates *SPL10* and *SPL11* to prevent premature entry into maturation during seed development (Willmann, Mehalick et al. 2011). While FUS3 directly induces expression of *pri-miRNA156 A* and *C* encoding genes (Wang and Perry, 2003 and Figure 2.6d), it has no effect on accumulation of *pri-miRNA156D* that is also not a direct target. Meanwhile, all three of these genes are associated with ABI3 and through the

developmental timecourse, ABI3 first induces these genes and then switches to lack of significant response (*A* and *C*) to repression (*D*) leading to increased accumulation of *SPL10/11* transcripts, supporting maturation programs (Figure 2.6). How might this switch occur? Part of the transition likely involves the timeframe of protein accumulation of FUS3 and ABI3 (*FUS3* transcript peaks at about 11-12 days and *ABI3* at 13-14 days; (Boulard, Fatihi et al. 2017). In addition, interactions may play a key role in determination of induction or repression. There are *cis* elements unique to the binding site of ABI3 to the gene encoding *MIR156D* that are not present for *MIR156A* or *C*. One such unique element is a pyrimidine box (TTTTTTCC, analysis of binding regions of FUS3/ABI3 performed with New Place (https://www.dna.affrc.go.jp/PLACE/?action=newplace). Pyrimidine boxes have been reported to be involved in ABA/GA regulation of late seed programs (Cercós et al., 1999).

Another interesting possibility differentiating *miRNA156A/C* from *D* has to do with the "passenger" strand of the *pri-gene miR156A, C* and *D*. The predicted mature *miRNA156* is identical for all three genes, however the complementary (also called passenger) strand of *D* is different from *A* and *C* with about 30% mismatch with the best alignment (miRBase, http://www.mirbase.org/help/targets.shtml; (Kozomara, Birgaoanu et al. 2019). Although the passenger strand has been considered a by-product of the mature miRNA in the past, data is accumulating to indicate that passenger strands may also have biological functions (reviewed in(Liu, Meng et al. 2017)). In birch (*Betula luminifera*), *miRNA156A** (* indicates the guide/passenger strand) is more abundant than *miRNA156A* (reviewed in (Liu, Meng et al. 2017)).

### 2.3.5 Seed Dormancy and Germination

A positive feedback loop exists between *MIR160*, ARF10, ARF16 and ABI3 to establish and maintain seed dormancy (Liu, Zhang et al. 2013) Figures 2.9 and 2.11). How might this loop be broken to allow completion of seed germination? MFT is a phosphatidylethanolamine-binding protein that promotes embryo growth, resulting in completion of germination (Xi, Liu et al. 2010). It is a direct repressed target of ABI3 (Xi, Liu et al. 2010); this manuscript), while ABI5 promotes induction of expression of *MFT*. ABI3 and ABI5 interact and both bind to a region approximately 1600 bp 5' of the start codon (*MFT-2* region; (Xi, Liu et al. 2010). ABI5, but not ABI3, binds to a region approximately 850 bp upstream of the start codon (*MFT-3* region; (Xi, Liu et al. 2010). Our results agree with a strong binding site overlapping with *MFT-2* and no binding peak associated with *MFT-3* (Figure 2.13). Two additional binding sites were found within the gene that were not tested in Xi et al. (Xi, Liu et al. 2010) (peaks 953 and 1889; Figure 2.13). Because ABI5 is a directly up-regulated target of ABI3, it is intriguing to consider that ABI3 (and ABI5) first bind to *MFT-2* to repress *MFT* expression. As ABI3 up-regulates *ABI5*, enough ABI5 protein may accumulate to bind to *MFT-3* and promote expression of *MFT*. MFT in turn directly represses *ABI5* and the binding region for MFT at ~1500 bp overlaps with that of ABI3 (Xi, Liu et al. 2010) Figure 2.14). MFT has not been shown to directly bind DNA. FD, a bZIP TF, is involved in recruiting the MFT homolog, FT. to *APETALA2* regulatory regions. Possibly a similar mechanism is in play for MFT. An intriguing possibility is that ABI5 and/or the antagonistic bZIP, EEL are involved (Bensmihen, Rippa

et al. 2002). *EEL* is also a directly induced target of ABI3 and EEL and ABI5 can interact (Ufaz, Shukla et al. 2011).

## 2.4 METHODS

### 2.4.1 Transgene Construct

For the *ABI3pro:ABI3-c-myc* and *ABI3pro:ABI3-HA* constructs, about 2.3 kb 5' of the start codon of *ABI3* and the entire genomic region excluding the stop codon (~5.2 kb total) was amplified and cloned into pENTR/D-TOPO vector (Invitrogen) following the manufacturer's direction. The insert was moved into the destination vectors pGWB19 (10x c-myc) or pGWB13 (3x HA) (Nakagawa, Ishiguro et al. 2009); obtained from Dr. T. Nakagawa, Shimane University) following the manufacturer's instruction for Gateway LR Clonase II Enzyme mix (Invitrogen). The transgene was introduced into *abi3-5* homozygous mutants by the floral dip method.

### 2.4.2 Plant Material

Arabidopsis Ler wild-type, and *abi3-5* plants were grown as described in (Wang and Perry 2013). For *abi3-5*, siliques from heterozygous plants were opened under sterile conditions at a relatively mature stage, *abi3-5* seeds identified by green color and moved to germination medium to establish the homozygous mutant.

To establish embryonic culture tissue (ECT) for ChIP experiments, a *35Spro:AGL15* transgene was introduced into the *abi3-5* background by transformation. Seeds were removed from siliques when nearly dry, put onto moist potting mix and allowed to develop two true leaves. Selection of transformants was performed by spraying with

kanamycin as described by Xiang et al. (Xiang, Han et al. 1999). Transformants were verified by PCR and used to cross with the *ABI3pro:ABI3-myc abi3-5* plants. Developing embryos were cultured to initiate ECT (Harding, Tang et al. (2003).

Developing zygotic embryos of green bent cotyledon stage were isolated for RNA extraction.

### 2.4.3 Expression Microarrays

Flowers of Ler WT and *abi3-5* were tagged on the day that they opened. Seeds were collected at 15-16 day after flowering (daf), and flash frozen in liquid nitrogen. RNA was extracted following the protocol of Chang et al. (Chang, Puryear et al. 1993). The RNA was further cleaned with a RNeasy Plant Minikit column. Three biological replicates were prepared and sent to the University of Kentucky Microarray Core Facility for probe generation and hybridization to the GeneChip™ Arabidopsis 1.0 ST array (ThermoFisher).

The ST arrays do not have absent, marginal or present calls associated with probes. To determine what genes were likely not expressed in our samples, we looked at the average of the intron controls on the array. This signal averaged 112. We also looked at the list of genes that were flagged as leaf specific (Zhang, Bishop et al. 2012) and found the signal from these genes was generally less than 100. The BioB control that has been suggested as a cutoff as present but low expressed had a signal of ~860 and AGL15 that should be present but low in older developing seeds was ~1000. We therefore removed any genes for which the signal was less than 300. Signals of 300-500 were highlighted red to indicate possible marginal expression.

### 2.4.4 ChIP-chip and Data Analysis

ChIP was performed as in (Wang and Perry 2013) except ECT expressing *ABI3pro:ABI3-myc* in an *abi3-5 (Ler)* background was used and tissue was treated with 10 µM ABA for 24 hour prior to treatment with formaldehyde. The control was the same tissue but ChIP performed without primary antibody. Probe generation and hybridization and data analysis were performed as in Zheng et al. (Zheng, Ren et al. 2009).

### 2.4.5 ChIP-qPCR and qRT-PCR

Experiments to confirm binding by ABI3 (ChIP-qPCR) and response to ABI3 accumulation (qRT-PCR) were performed as described in (Wang and Perry 2013). For follow-up ChIP experiments, ECT expressing *ABI3pro:ABI3-HA* was used with anti-HA (Biolegend, Rabbit polyclonal 9023), or developing siliques expressing *ABI3pro:ABI3-c-myc* were used. The specific primers used for these experiments are listed in Table 2.7.

The ChIP-chip and expression array data have been deposited in the National Center for Biotechnology Information's Gene Expression Omnibus and is accessible through GEO series accession number GSE150561. Arabidopsis Genome Initiative (AGI) name and full name of genes referenced in this manuscript are provided in Table 2.8.

**Table 2.1** Location of sequences co-immunoprecipitated with ABI3-c-MYC

Sequences were obtained and mapped to the Arabidopsis genome using CisGenome. Some totals are different from sums of regions because some sequences span more than one region. TES, transcriptional end site; TSS, transcriptional start site; up/down 1 kb, within 1 kb of the TSS and/or TES; UTR, untranslated region of the transcript 5' and 3' of the coding sequences (CDS).

| **Region of gene** | **ABI3 bound sites (% of 2510)** | **ABI3 bound and induced (% of 317)** | **ABI3 bound and repressed (% of 87)** |
|---|---|---|---|
| **TSS up 1 kb** | 1190 (47%) | 180 (57%) | 38 (44%) |
| **TES down 1 kb** | 717 (29%) | 100 (32%) | 25 (29%) |
| ***Total intergenic (1kb)*** | 1859 (74%) | 188 (59%) | 69 (79%) |
| **5' UTR** | 140 (6%) | 34 (11%) | 7 (8%) |
| **CDS** | 297 (12%) | 63 (20%) | 1 (1%) |
| **Intron** | 149 (6%) | 30 (9%) | 9 (10%) |
| **3' UTR** | 90 (4%) | 5 (2%) | 1 (1%) |
| ***Total intragenic*** | 651 (26%) | 129 (41%) | 18 (21%) |

**Table 2.2** *Cis* motifs present in peaks bound by ABI3. *, significant difference from the matched sets as assessed by chi-squared tests. na, not assessed.

| *Cis* motifs | All bound sites (% of 2510) | Bound and ABI3 induced (% of 316) | Bound and ABI3 repressed (% of 87) |
|---|---|---|---|
| **Long RY (CATGCATG)** | 162 (6%)* | 41 (13%)* | 5 (6%) |
| **Short RY (CATGCA)** | 999 (40%)* | 187 (59%)* | 20 (23%) |
| **G-box** | 1286 (51%)* | 187 (59%)* | 38 (44%)* |
| **E-box (CANNTG)** | na | 304 (96%)* | 81 (93%) |
| **CACCG(ABI4)** | na | 219 (69%)* | 31 (36%) |
| **Average of three matched sets** | 2% (full RY)<br>26% (part. RY)<br>11% (G box) | 3% (long RY)<br>39% (part. RY)<br>14% (G box)<br>98% (E box)<br>61% (CACCG) | 1% (long RY)<br>24% (part. RY)<br>5% (G box)<br>95%(E box)<br>34% (CACCG) |
| **Average width** | 829 bp | 1368 bp | 681 bp |

**Table 2.3** Select previously identified ABI3 targets and presence in ABI3 regulon experiments

ABI3 peak#, this manuscript; FUS3 peak#, (Wang and Perry 2013); Mönke, ChIP and transcriptome data from (Mönke, Seifert et al. 2012); Bargmann, data from (Bargmann, Marshall-Colon et al. 2013). Sources of the ABI3 targets are shown in the footnote. N: not identified; Y: also recovered.

| Target gene | Tissue/ Experiment | ABI3 peak# (FUS3 peak#) | Mönke | | Bargmann |
|---|---|---|---|---|---|
| | | | ChIP | transcript | TARGET |
| *SOM;At1g03790* | Seeds/ChIP[1,2] | 66 | N | Y | Y |
| *At4g36700* | | 71 (222) | N | Y | N |
| *At5g22470* | | 59 | Y | Y | Y |
| *At5g62800* | | 58 (121) | Y | N | Y |
| *MFT;At1g18100* | Seeds/ChIP[3] | 953, 1889 and 149 | N | N | N |
| *RD29B;At5g52300* | Protoplasts/ Transactivation[4] | 136 | N | N | Y |
| *RD29A;At5g52310* | | 823 | N | N | Y |
| *SDH2.3;At5g65165* | Protoplasts/ Transactivation[5] | 41 (1287) | N | Y | N |
| *OLE1;At4g25140* | Embryo, protoplasts/ Reporter, EMSA[6] | 5 (198) | Y | Y | Y |
| *HSFA9;At5g54070* | Protoplasts/ Transactivation[7] | 388 | N | N | N |
| *CRU1;At5g44120* | Dehydration stress in seedlings/ChIP[8] | 82 | Y | Y | Y |
| *CRU3;At4g28520* | | 83 | Y | Y | N |
| *LEA76;AT3g15670* | | 18 | Y | Y | Y |
| *At2g21490* | | 77 | Y | Y | N |
| *At3g02480* | | 33 | Y | N | Y |
| *At2g40170* | | 163 | N | N | Y |
| *ABI3;At3g24650* | | 84 | N | N | N |

[1]Lim et al., 2013; [2] Park et al., 2011; [3]Xi et al., 2010; [4]Nakashima et al., 2006; [5]Restovic et al., 2017; [6]Baud et al., 2016; [7]Kotak et al., 2007; [8]Bedi et al., 2016

**Table 2.4** Genes that are potentially direct targets of LEC2, FUS3 and ABI3

Data for FUS3 is from (Wang and Perry 2013) and for LEC2-GR from (Braybrook, Stone et al. 2006). daf, days after flowering

| AGI | TAIR 10 description | ABI3 peak | *abi3/Ler* 15-16 daf | FUS3 peak | *fus3*/Col WT | | Induced by LEC2-GR |
|---|---|---|---|---|---|---|---|
| | | | | | 8 daf | 12 daf | |
| *AT1G67100* | LOB domain-containing protein 40 | 51 | 0.475 | 2 | 0.535 | 0.430 | 1, 4 h |
| *AT2G34700* | Pollen Ole e 1 allergen and extensin family protein | 436 | 0.153 | 4 | 0.135 | 0.048 | 1, 4 h |
| *AT3G54940* | Papain family cysteine protease | 2 | 0.053 | 82 | 0.826 | 0.343 | 1, 4 h |
| *AT1G65090* | Nucleolin | 35 | 0.359 | 108 | 0.227 | 0.234 | 4 h |
| *AT4G25140* | Oleosin 1 | 5 | 0.672 | 198 | 0.344 | 0.152 | 4 h |
| *AT4G36700* | RmlC-like cupins superfamily protein | 71 | 0.480 | 222 | 0.415 | 0.104 | 4 h |
| *AT4G01460* | basic helix-loop-helix (bHLH) DNA-binding | 2227 | 1.603 | 313 | 0.463 | 0.860 | 1, 4 h |
| *AT2G41070* | Basic-leucine zipper (bZIP) transcription factor | 104 | 0.284 | 412 | 0.367 | 0.356 | 1, 4 h |
| *AT5G60760* | P-loop containing nucleoside triphosphate hydrolases | 265 | 0.176 | 458 | 0.868 | 0.209 | 1, 4 h |
| *AT5G45830* | delay of germination 1 (DOG1) | 341 | 0.403 | 567 | 0.221 | 0.041 | 1, 4 h |
| *AT5G40420* | Oleosin 2 | 47 | 0.395 | 704 | 0.343 | 0.152 | 4 h |
| *AT5G54740* | Seed storage albumin 5 | 49 | 0.199 | 878 | 0.100 | 0.063 | 1, 4 h |
| *AT1G01720* | NAC domain transcriptional regulator | 107 | 0.207 | 1008 | 0.512 | 0.244 | 1 h |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| *ABI3 repressed* | *Direct (87)* | 19 (21.8%) | 4 (4.6%) | 22 (25.3%) | 10 (11.5%) | 21 (24.1%) | 10 (11.5%) | 2 (2.3%) | 1 (1.1%) |
| | *Indirect (797)* | 17 (2.1%) | 36 (4.5%) | 37 (4.6%) | 73 (9.2%) | 63 (7.9%) | 174 (21.8%) | 15 (1.9%) | 11 (1.4%) |

**Table 2.5** Comparison of ABI3 regulated genes to those involved in dormancy

Data for Ler Dormant, D versus Afterripened, AR genes from (Carrera, Holman et al. 2007).

| **Comparison** | | **Ler D up vs Ler AR (1595 genes)** | **Ler AR up vs Ler D (1945 genes)** |
|---|---|---|---|
| *ABI3 induced* | All (627) | 284; 45% | 10; 2% |
| | Direct (316) | 177; 56% | 3; 1% |
| *ABI3 repressed* | All (884) | 37; 4% | 244; 28% |
| | Direct (87) | 7; 8% | 28; 32% |

**Table 2.6.** Overlap between ABI3 direct and indirect targets (from this study) and ABA regulated genes during re-establishment of desiccation tolerance (Costa, Righetti et al. 2015). All comparisons are to 0 hr ABA treatment with the exception of 72 hr that is compared to 24 hr ABA treatment.

| | | **2 hr ABA** | | **12 hr ABA** | | **24 hr ABA** | | **72 hr ABA** | |
|---|---|---|---|---|---|---|---|---|---|
| | | Up | Down | Up | Down | Up | Down | Up | Down |
| **Number** | | 803 | 1449 | 1417 | 1842 | 2172 | 2785 | 504 | 346 |
| *ABI3 induced* | *Direct (317)* | 178 (56.2%) | 12 (3.8%) | 188 (59.3%) | 17 (5.4%) | 198 (62.5%) | 18 (5.6%) | 3 (0.9%) | 26 (8.2%) |
| | *Indirect (311)* | 48 (15.4%) | 19 (6.1%) | 68 (21.9%) | 23 (7.4%) | 84 (27.0%) | 24 (7.7%) | 2 (2.3%) | 1 (1.1%) |
| *ABI3 repressed* | *Direct (87)* | 19 (21.8%) | 4 (4.6%) | 22 (25.3%) | 10 (11.5%) | 21 (24.1%) | 10 (11.5%) | 2 (2.3%) | 1 (1.1%) |
| | *Indirect (797)* | 17 (2.1%) | 36 (4.5%) | 37 (4.6%) | 73 (9.2%) | 63 (7.9%) | 174 (21.8%) | 15 (1.9%) | 11 (1.4%) |

**Table 2.7:** Primers used in this study presented 5'→3'

| AGI | Gene name | Forward | Reverse |
|---|---|---|---|
| **For cloning** | | | |
| *ABI3pro:ABI3* | *At3g24650* | CACCGATACGTGTACGTTTAGGTGGC | TTTAACAGTTTGAGAAGTTGGTGAAGC |
| **For qRT-PCR** | | | |
| *miRNA 160B* | *At4g17788* | TGCCTGGCTCCCTGTAT | TTGACTACTCTGTACGCCACTA |
| *ARF10* | *At2g28350* | CACACTCGATCTCTCGGTTATTG | GAGTCAACAAATCCGACCTCTCTTC |
| *ARF16* | *At4g30080* | ACAGGGCATTGTAAAGTGTTTATG | GAGTTTCCGACTCAATTCTTCGTATG |
| *ABI3, set 1* | *At3g24650* | TCAGCTTCTGCTATGCCACGTCA | TTACCCACGTCGCTTTGCTTCA |
| *ABI3, set 2* | *At3g24650* | ACGCATCAGGAAACTGTGACGA | TCAATGCAATCACCACCGCCTT |
| *miRNA 156A* | *At2g25095* | AGAAAGAGCAGTGAGCACGCAA | AGAGAGTGAGCACACAAAGGCA |
| *miRNA 156C* | *At4g31877* | AGCCCTGCAGTATTCATCTCT | CCCATCTCTCACTTTCTCTCTCA |
| *miRNA 156D* | *At5g10945* | GATGGGGGAAAAGAAGTTGACAG | GAGCACGCAAAAGCAACCATATAC |
| *SPL10* | *At1g27370* | AGTTGTTGTGAGTGGCCTGGAA | AGAAAGACGTTTGCGGCAGCTT |
| *SPL11* | *At1g27360* | AAAGTTAGCGTGAGTGGCCTGGAA | TCGGCAGCTTCGTTTCTTCTCA |
| *TUA3* | *At5g19770* | TGGTGCCCAACTGGGTTCAAAT | ACCTCTGCAACTGCTGTGTTGT |
| **For ChIP-qPCR** | | | |
| *miRNA 156D* | *At5g10945* | TCCTTAATCACTCACTCTCACATCT | CGCTATATAGAGAGAGAAGTAAAGC |
| *miRNA 160B* | *At4g17788* | CACAACACATTTGTCTCGCTATTC | CTGTCACCTAATCAGTGTCGATG |
| *Intergenic region* | Non-bound region | GAACTACTCGGTTTGCGAATTG | CCTTGCCGATCCTGATGAATA |

**Table 2.8:** AGI and full names or descriptions of genes/proteins.

| Name | AGI | Full name |
|---|---|---|
| --- | At4g36700 | RmlC-like cupins superfamily protein |
| --- | At5g62800 | Protein with RING/U-box and TRAF-like domain |
| --- | At2g21490 | Dehydrin LATE EMBRYOGENESIS ABUNDANT |
| --- | At3g02480 | LATE EMBRYOGENESIS ABUNDANT |
| --- | AT2G34700 | Pollen Ole e 1 allergen and extensin family protein |
| --- | AT3G54940 | Papain family cysteine protease |
| --- | AT1G65090 | Nucleolin |
| --- | AT4G01460 | basic helix-loop-helix (bHLH) DNA-binding |
| --- | AT5G60760 | P-loop containing nucleoside triphosphate hydrolases |
| --- | At3g53040 | LATE EMBRYOGENESIS ABUNDANT |
| --- | At1g15330 | Cystathionine β-synthase family protein |
| ABA1 | AT5G67030 | ABSCISIC ACID DEFICIENT1 |
| ABCG31 | AT2G29940 | ATP-BINDING CASSETTE G31 |
| ABI1 | AT4G26080 | ABSCISIC ACID INSENSITIVE1 |
| ABI3 | AT3G24650 | ABSCISIC ACID INSENSITIVE3 |
| ABI4 | AT2G40220 | ABSCISIC ACID INSENSITIVE4 |
| ABI5 | AT2G36270 | ABSCISIC ACID INSENSITIVE5 |
| ACP1 | AT3G05020 | ACYL CARRIER PROTEIN1 |
| AGL15 | AT5G13790 | AGAMOUS-LIKE15 |
| AGL67 | AT1G77950 | AGAMOUS-LIKE67 |
| ANAC032 | AT1G77450 | ARABIDOPSIS NAC DOMAIN CONTAINING PROTEIN32 |
| ANAC102 | AT5G63790 | ARABIDOPSIS NAC DOMAIN CONTAINING PROTEIN102 |
| ANAC2/ATAF1 | AT101720 | ARABIDOPSIS NAC DOMAIN CONTAINING PROTEIN2 |
| AP2 | AT4G36920 | APETALA2 |
| ARF10 | AT2G28350 | AUXIN RESPONSE FACTOR10 |
| ARF16 | AT4G30080 | AUXIN RESPONSE FACTOR16 |
| ATEM1 | AT3G51810 | EARLY-METHIONINE LABELLED1 |
| ATEM6 | AT2G40170 | EARLY-METHIONINE LABELLED6 |
| ATTIL | AT5G58070 | TEMPERATURE-INDUCED LIPOCALIN |
| BBM | AT1G17430 | BABY BOOM |
| CAP160 | At4g25580 | COLD ACCLIMATION PROTEIN 160 |
| CBSX4 | AT1G80090 | CYSTATHIONINE β-SYNTHASE FAMILY |
| CRU1 | AT5G44120 | CRUCIFERIN A |
| CRU3 | AT4G28520 | CRUCIFERIN 3 |
| CYP707A2 | AT2G29090 | CYTOCHROME P450, FAMILY 707, SUBFAMILY A2 |
| DOG1 | AT5G45830 | DELAY OF GERMINATION1 |
| DREB2A | AT5G05410 | DRE-BINDING PROTEIN2A |
| ECP63 | AT2G36640 | EMBYRONIC CELL PROTEIN63 |
| EEL | AT2G41070 | ENHANCED EM LEVEL |

**Table 2.8** (continued)

| | | |
|---|---|---|
| EF1-α | AT5G60390 | ELONGATION FACTOR 1-α |
| FAD3 | AT2G29980 | FATTY ACID DESATURASE3 |
| FBH3 | AT1G51140 | FLOWERING BHLH3 |
| FUS3 | AT3G26790 | FUSCA3 |
| GAI | AT1G14920 | GIBBERELLIC ACID INSENSITIVE |
| GBF2 | AT4G01120 | G-BOX BINDING FACTOR2 |
| GBF3 | AT2G46270 | G-BOX BINDING FACTOR3 |
| GOLS2 | AT1g56600 | GALACTINOL SYNTHASE2 |
| HB5 | AT5G65310 | HOMEOBOX PROTEIN5 |
| HB7 | AT2G46680 | HOMEOBOX PROTEIN7 |
| HON/HAI2 | AT1G07430 | HONSU/HIGHLY ABA INDUCED PP2C PROTEIN POLYPEPTIDE2 |
| HSFA9 | AT5G54070 | HEAT SHOCK TRANSCRIPTION FACTOR A9 |
| KCS18/FAE1 | AT4G34520 | 3-KETO-CoA SYNTHASE18/FATTY ACID ELONGATION1 |
| LACS4 | AT4G23850 | LONG-CHAIN ACYL-CoA SYNTHETASE4 |
| LBD40 | AT1G67100 | LOB domain-containing protein 40 |
| LEA14 | AT1G01470 | LATE EMBRYOGENESIS ABUNDANT14 |
| LEA76 | AT3G15670 | LATE EMBRYOGENESIS ABUNDANT76 |
| LEC1 | AT1G21970 | LEAFY COTYELDON1 |
| LEC2 | At1G28300 | LEAFY COTYLEDON2 |
| MFT | AT1G18100 | MOTHER OF FT AND TFL1 |
| MIR156A | At2g25095 | MICRORNA 156A |
| MIR156C | At4g31877 | MICRORNA 156C |
| MIR156D | At5g10945 | MICRORNA 156D |
| MIR160B | AT4G17788 | MICRORNA 160B |
| NCED9 | AT1G78390 | NINE-CIS-EPOXY CAROTENOID DEOXYGENASE9 |
| NOL | AT5G04900 | NYC1-LIKE |
| NYC1 | AT4G13250 | NON-YELLOW COLORING1 |
| OLE1 | AT4G25140 | OLEOSIN1 |
| OLE2 | AT5G40420 | OLEOSIN2 |
| PARP3 | At5g22470 | POLY(ADP-RIBOSE) POLYMERASE3 |
| PIF1/PIL5 | AT2G20180 | PHYTOCHROME INTERACING FACTOR1/PHYTOCHROME INTERACTING FACTOR3-LIKE 5 |
| PIMT2 | AT5G50240 | PROTEIN-L-ISOASPARTATE METHYLTRANSFERASE2 |
| PLATZ1 | AT1G21000 | PLATZ TRANSCRIPTION FACTOR FAMILY |
| PLATZ2 | AT1G76590 | PLATZ TRANSCRIPTION FACTOR FAMILY |
| PP2A | AT1G69960 | SERINE/THREONING PROTEIN PHOSPHATASE 2A |
| RAB18 | AT5G66400 | RESPONSIVE TO ABA18 |
| RD26 | AT4G27410 | RESPONSIVE TO DESICCATION26 |
| RD29A/LTI78 | AT5G52310 | RESPONSIVE TO DESICCATION29A/LOW-TEMPERATURE-INDUCED78 |
| RD29B/LTI65 | AT5G52300 | RESPONSIVE TO DESICCATION29b/LOW-TEMPERATURE-INDUCED65 |

**Table 2.8** (continued)

| | | |
|---|---|---|
| RDO5 | AT4G11040 | REDUCED DORMANCY5 |
| RGA1 | AT2G01570 | REPRESSOR OF GA1-3 1 |
| RGL2 | AT3G03450 | RGA-LIKE2 |
| RGL3 | AT5G17490 | RGA-LIKE3 |
| SDH2.3 | AT5G65165 | SUCCINATE DEHYDROGENASE2-3 |
| SESA5 | AT5G54740 | SEED STORAGE ALBUMIN5 |
| SGR2 | AT4G11910 | STAY GREEN2 |
| SIP2 | AT3G57520 | SEED IMBIBITION2 |
| SNRK2.2 | AT3G50500 | SNF1-RELATED PROTEIN KINASE2.2 |
| SNRK2.3 | AT5G66880 | SNF1-RELATED PROTEIN KINASE2.3 |
| SNRK2.4 | AT1G10940 | SNF1-RELATED PROTEIN KINASE2.4 |
| SOM | AT1G03790 | SOMNUS |
| SPL10 | At1g27370 | SQUAMOSA PROMOTER BINDING PROTEIN-LIKE 10 |
| SPL11 | At1g27360 | SQUAMOSA PROMOTER BINDING PROTEIN-LIKE 11 |
| SRK2I | AT5G66880 | SUCROSE NONFERMENTING1 (SNF1)-RELATED PROTEIN KINASE |
| SSI2 | AT2G43710 | SUPPRESSOR OF SA INSENSITIVE2 |
| STS | AT4G01970 | STACHYOSE SYNTHASE |
| TIP3;1 | AT1G73190 | TONOPLAST INTRINSIC PROTEIN3;1 |
| TUA3 | AT5G19770 | TUBULIN $\alpha$3 |
| TUB2 | AT5G62690 | TUBULIN $\beta$ CHAIN2 |
| VAL1/HSI2 | AT2G30470 | VP1/ABI3-LIKE1/HIGH-LEVEL EXPRESSION OF SUGAR INDUCIBLE2 |
| VAL2/HSL | AT4G32010 | VP1/ABI3-LIKE2/HSI2-LIKE1 |
| VAL3 | AT4G21550 | VP1/ABI3-LIKE3 |
| VTE1 | AT4G32770 | VITAMIN E DEFICIENT1 |
| WRI1 | AT3G54320 | WRINKLED1 |
| XERO1 | AT3G50980 | DEHYDRIN XERO1 |

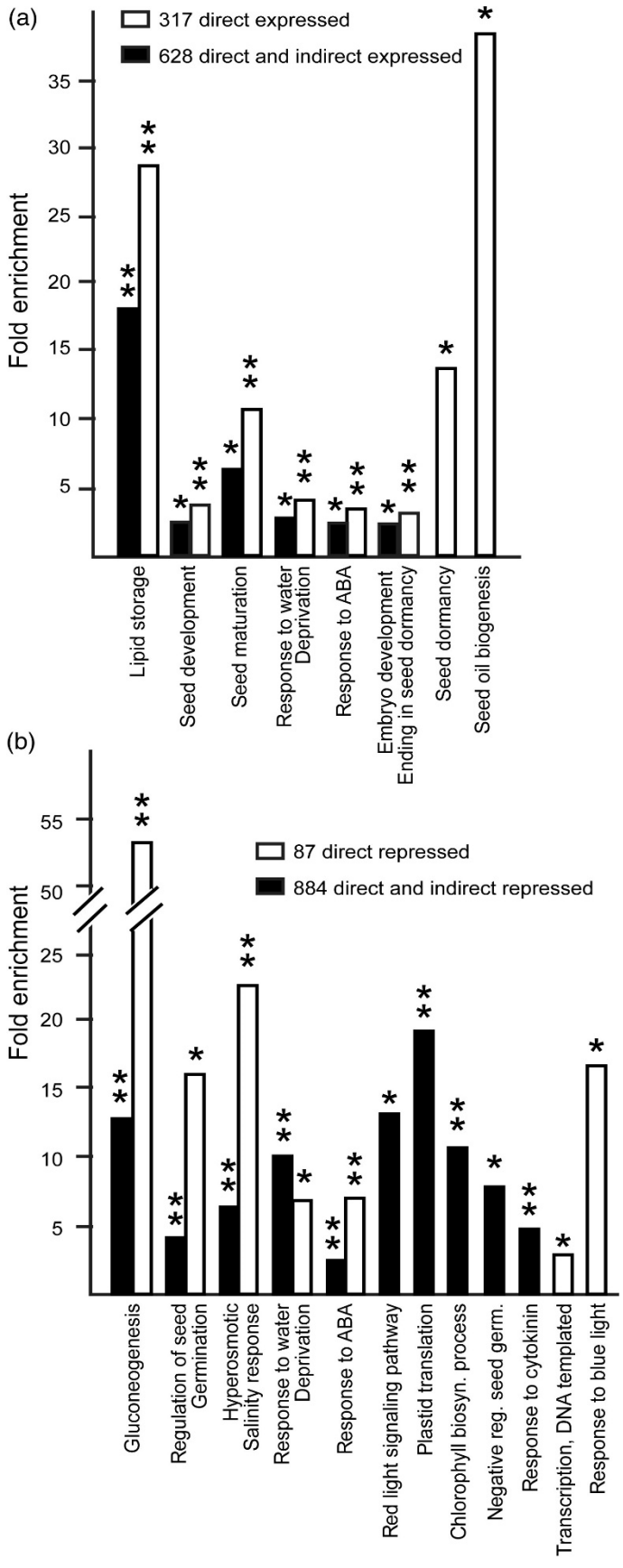
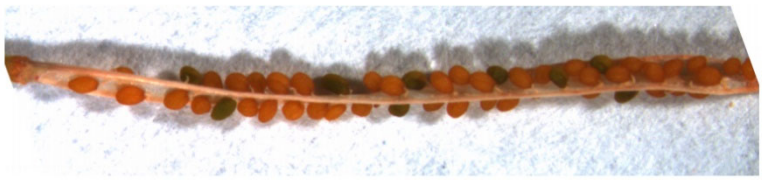

(a)
317 direct expressed
628 direct and indirect expressed
Fold enrichment
35
30
25
20
15
10
5
Lipid storage
Seed development
Seed maturation
Response to water Deprivation
Response to ABA
Embryo development Ending in seed dormancy
Seed dormancy
Seed oil biogenesis
(b)
87 direct repressed
884 direct and indirect repressed
Fold enrichment
55
50
25
20
15
10
5
Gluconeogenesis
Regulation of seed Germination
Hyperosmotic Salinity response
Response to water Deprivation
Response to ABA
Red light signaling pathway
Plastid translation
Chlorophyll biosyn. process
Negative reg. seed germ.
Response to cytokinin
Transcription, DNA templated
Response to blue light

**Figure 2.1** Functional categorization of ABI3 direct and indirect responsive target genes using the PANTHER classification system statistical overrepresentation test with default settings. Select GO biological processes are shown as fold enrichment over the Arabidopsis genome.

(a) Expressed in response to ABI3.

(b) Repressed in response to ABI3. * FDR<0.05; ** FDR<0.01.

*1609-13-2: pABI3:gABI3-HA plant 5*



*1609-19-2: pABI3:gABI3-MYC plant 1 (top) and 2 (bottom).*

**Figure 2.2** Complementation of *abi3-5* with the epitope tagged transgenes driven by the native promoter. The transgenes in 1609-13-2-5 and 1609-19-2-2 complemented the green seed phenotype of *abi3-5*, but 1609-19-2-1 did not.

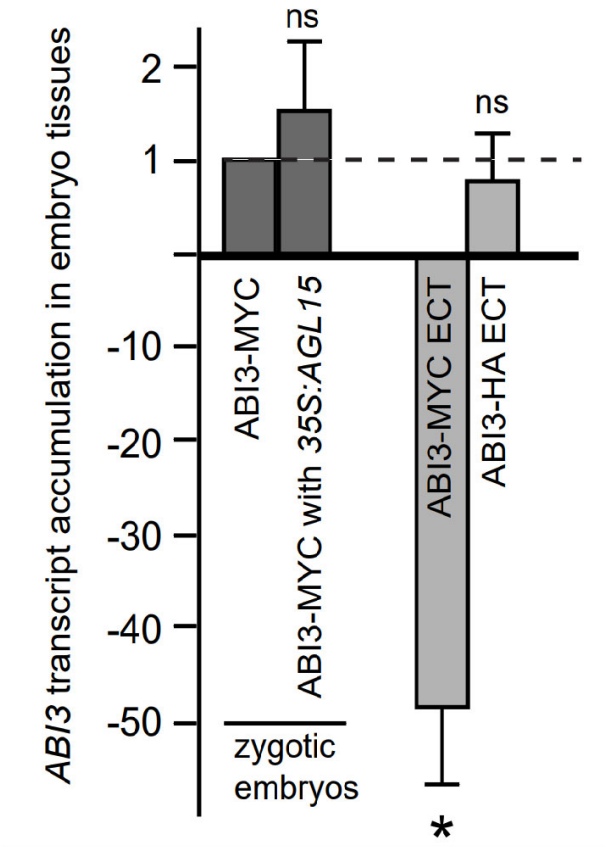
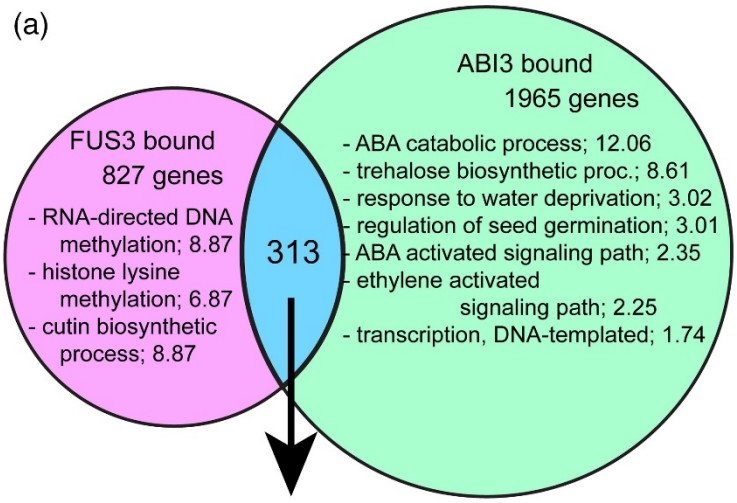



**Figure 2.3** ECT used for ChIP-seq does not overaccumulate *ABI3* transcript. Transcript accumulation from endogenous *ABI3* and *ABI3pro:ABI3* transgenes in the presence and absence of the *35Spro:AGL15* transgene. *ABI3* transcript from isolated green bent cotyledon zygotic embryos from *ABI3pro:ABI3* was set to one and other zygotic and somatic embryo tissues compared. *significant at P<0.01; ns, not signficant.

(a)




ABI3 and FUS3 bound
GO Categories enriched for:

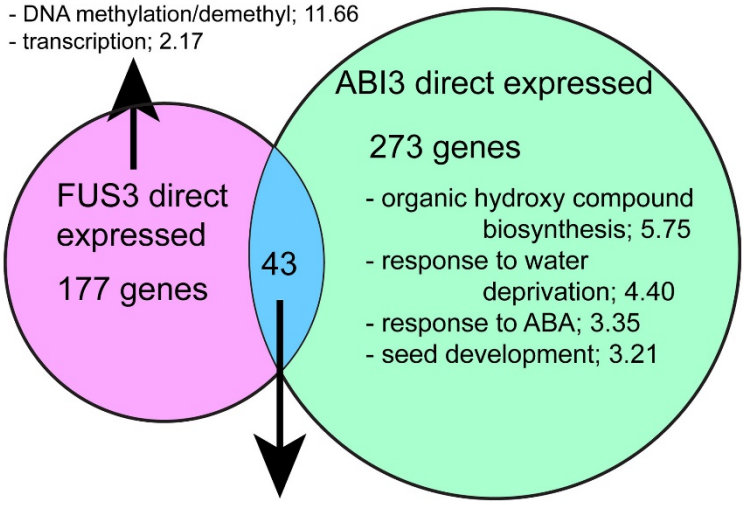
Genes with distinct binding sites for FUS3 and ABI3 (122 genes):
- response to ethylene; 7.00
- response to water deprivation; 5.62

Genes with overlapping binding sites for FUS3 and ABI3 (219 genes):
- lipid storage; 37.75
- seed oil biogenesis; 77.04
- seed dormancy process; 20.74

Shared between genes with distinct and overlapping sites for FUS3 and ABI3:
- seed development; 4.11
- response to karrikin; 8.05
- transcription, DNA-templated; 2.87
- response to hormone; 2.7

(b)

- somatic embryogenesis; 35.66
- seed development; 3.97
- lipid storage; 24.72
- reg. of seed maturation; 23.18
- cutin biosyn. proc.; 23.18
- positive reg. seed germination; 22.07
- DNA methylation/demethyl; 11.66
- transcription; 2.17




ABI3 and FUS3 direct expressed
- lipid storage; >100
- nutrient reserve activity; 39.09
- lipid droplet; 93.54

**Figure 2.4** Genes associated with shared a**n**d unique targets of ABI3 and FUS3.

(a) PANTHER classification system was used to discover significantly (FDR<0.05) overrepresented categories of genes for which regulatory regions are bound by both ABI3 and FUS3, or that are unique for one of these factors. (b) As in (a) but for genes directly expressed by these factors. Select Biological Process categories; Fold Enrichment compared to the Arabidopsis genome are shown.

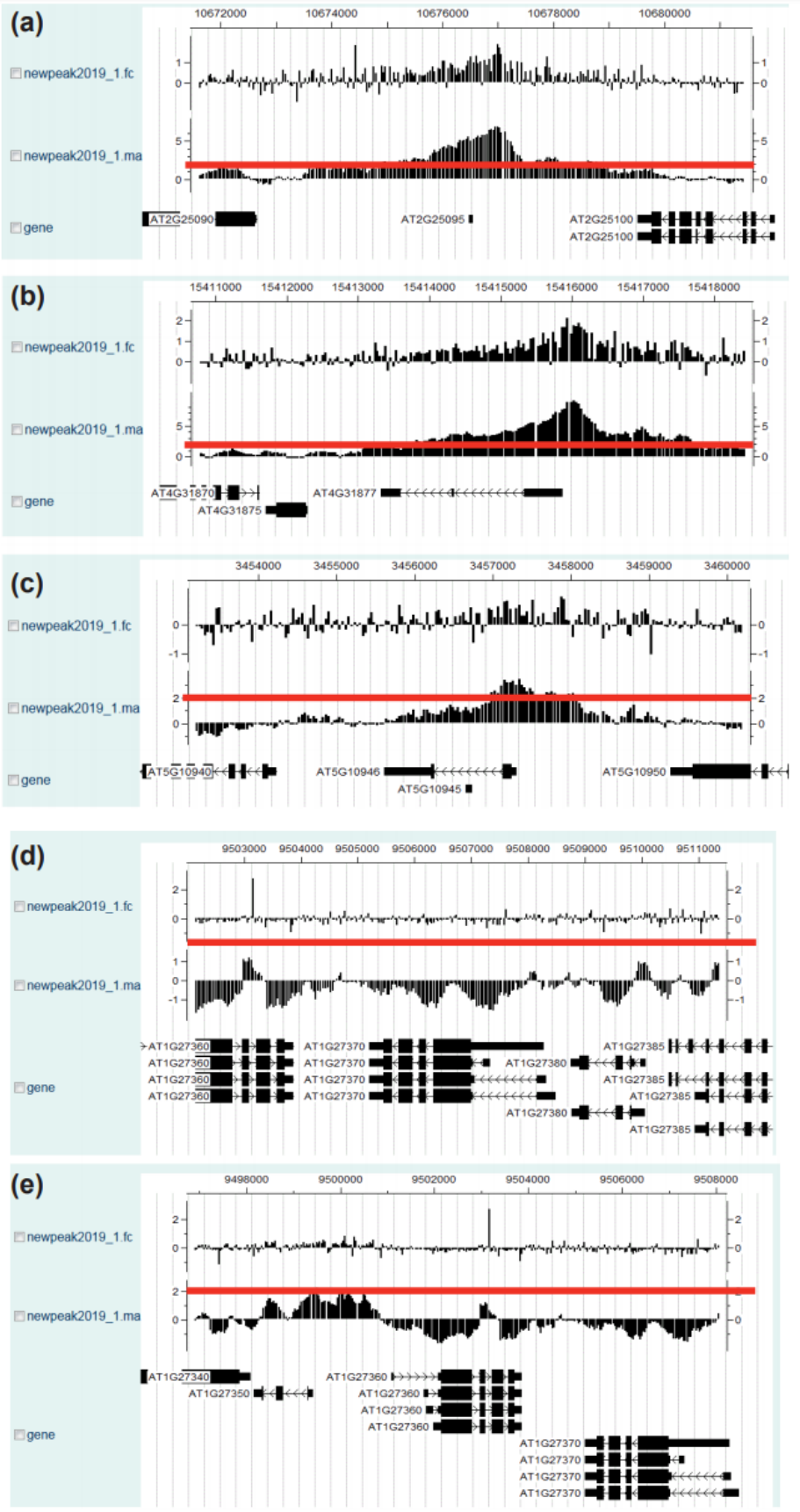

(a)
10672000
10674000
10676000
10678000
10680000
newpeak2019_1.fc
newpeak2019_1.ma
gene
AT2G25090
AT2G25095
AT2G25100
AT2G25100
(b)
15411000
15412000
15413000
15414000
15415000
15416000
15417000
15418000
newpeak2019_1.fc
newpeak2019_1.ma
gene
AT4G31870
AT4G31875
AT4G31877
(c)
3454000
3455000
3456000
3457000
3458000
3459000
3460000
newpeak2019_1.fc
newpeak2019_1.ma
gene
AT5G10940
AT5G10946
AT5G10945
AT5G10950
(d)
9503000
9504000
9505000
9506000
9507000
9508000
9509000
9510000
9511000
newpeak2019_1.fc
newpeak2019_1.ma
gene
AT1G27360
AT1G27370
AT1G27380
AT1G27385
(e)
9498000
9500000
9502000
9504000
9506000
9508000
newpeak2019_1.fc
newpeak2019_1.ma
gene
AT1G27340
AT1G27350
AT1G27360
AT1G27370

**Figure 2.5** Cisgenome traces showing potential ABI3 binding to regulatory regions associated with *At2g25095, miR156A* (a), *At4g31877, miR156C* (b), *At5g10945, miR156D* (c), *At1g27370, SPL10* (d) and *At1g27360, SPL11* (e). The red line indicates a 2-fold moving average comparing DNA recovered by immune ChIP to no antibody control.

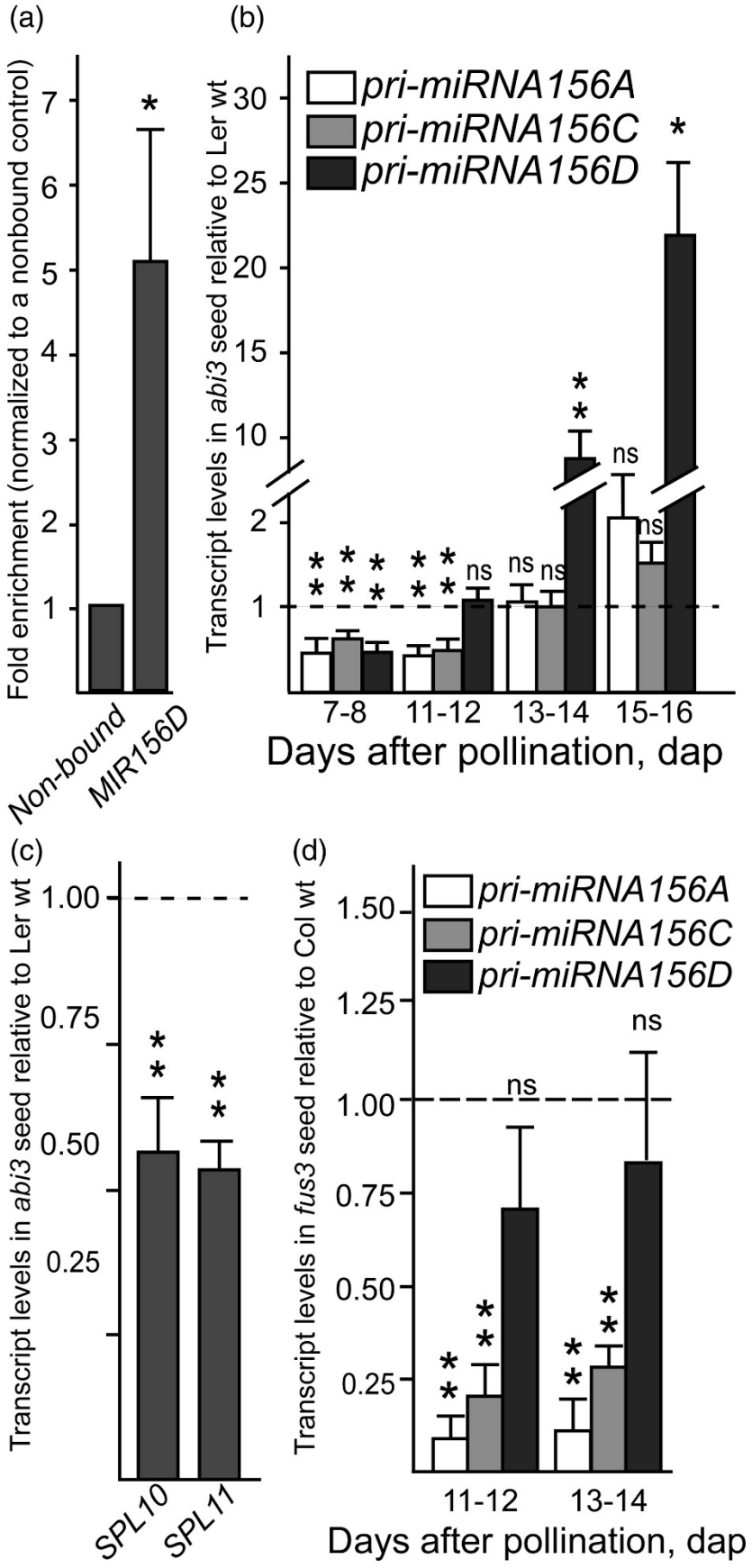

(a)
(b)
(c)
(d)
Fold enrichment (normalized to a nonbound control)
Non-bound
MIR156D
pri-miRNA156A
pri-miRNA156C
pri-miRNA156D
Transcript levels in abi3 seed relative to Ler wt
7-8
11-12
13-14
15-16
Days after pollination, dap
ns
Transcript levels in abi3 seed relative to Ler wt
SPL10
SPL11
Transcript levels in fus3 seed relative to Col wt
11-12
13-14
Days after pollination, dap

**Figure 2.6** The miRNA-encoding gene MIR156D is a directly expressed ABI3 target at earlier stages of seed development, but repressed at later stages of seed development. (a) Fold enrichment of regulatory regions associated with MIR156D from ChIP using HA antibody and ABI3-HA tissue. Data are normalized to a non-bound region of the genome. (b) Transcript accumulation corresponding to pri-miRNA156A, -C and -D in *abi3-5* developing seeds relative to Ler wild-type (wt) set to 1 (dotted line). (c) Transcript accumulation of the miRNA156 target genes SPL10 and SPL11 in *abi3-5* compared with Ler wt at 15–16 dap. (d) Transcript accumulation as in (b) but comparing *fus3-3* to Col wt set to 1 (dotted line). Means and SEMs are shown for three to five biological replicates. *Significant at $P < 0.05$; **Significant at $P < 0.01$.

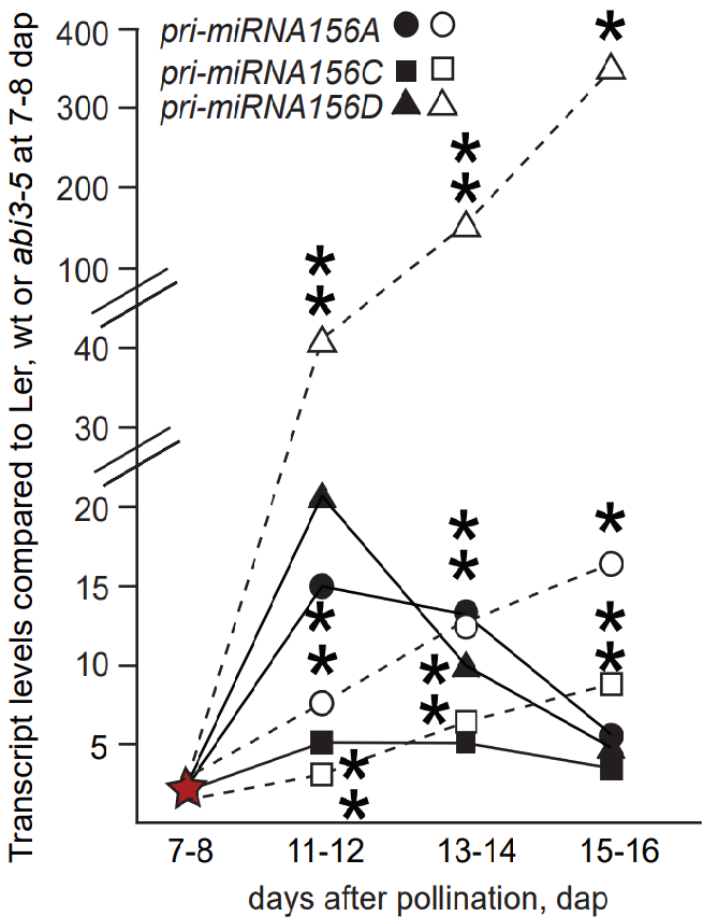



**Figure 2.7** Timecourse of transcript accumulation of *pri-miRNA156A, C* and *D*. Transcript accumulation at 7-8 dap is set to 1. Other timepoints for Ler and for *abi3-5* are compared to the same genotype at 7-8 dap. Filled symbols with solid lines represent data for Ler, WT while open symbols with dotted lines represent the *abi3-5* data. Results shown are for one experiment that is representative of two other biological repeats. Please note the breaks in the y-axis. *significant at P<0.05; **significant at P<0.01.

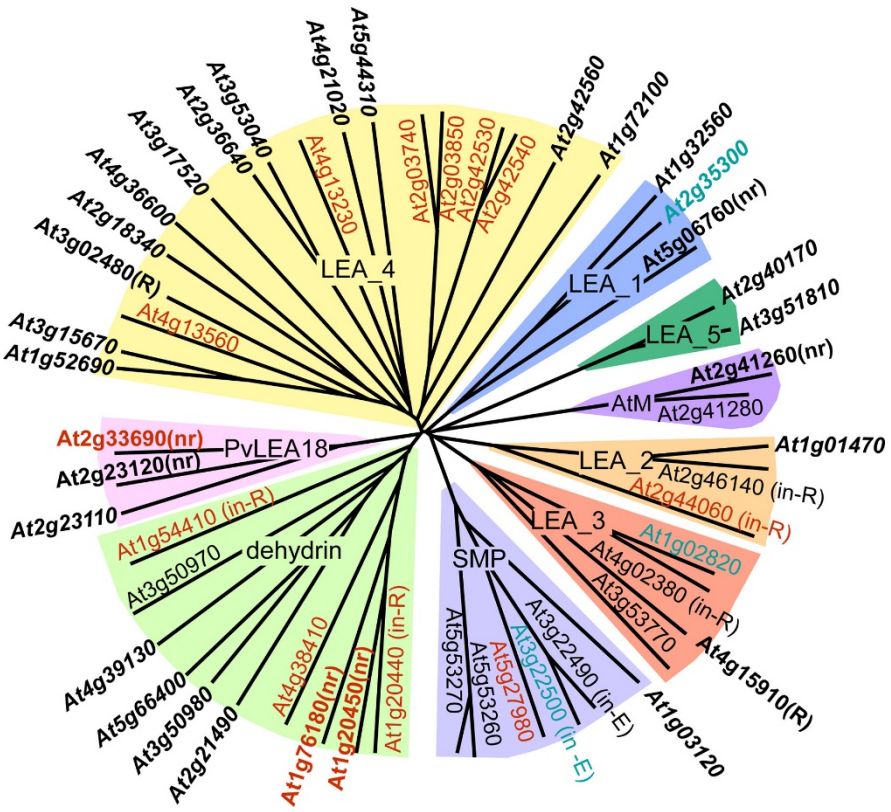



**Figure 2.8** *LEA* genes regulated by ABI3. Direct responsive genes are indicated by the AGI at the end of the line on the dendrogram with ABI3 direct expressed genes shown in bold italics. An (R) following the AGI indicates the gene is directly repressed by ABI3. Placement partially on the line with (nr) indicates the gene shows direct association with ABI3 but no significant response at P<0.05. Black text indicates the gene has an expression pattern that includes seeds. Red text indicates that the gene is not expressed in seeds. Blue text indicates expression only in response to stress or salt. Text on the lines indicates no detected binding by ABI3 in the ChIP-chip study. in-R, the gene appears indirectly repressed by ABI3; in-E, the gene appears indirectly expressed by ABI3. The figure is a modification of (Hundertmark and Hincha 2008).

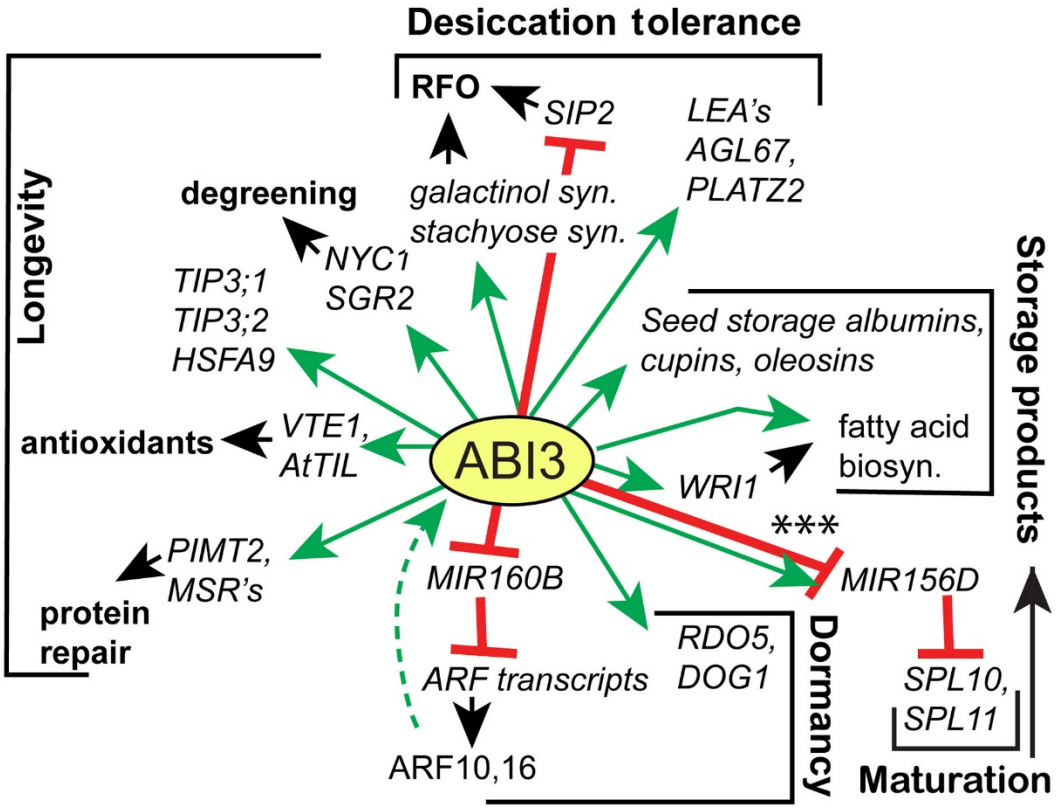



**Figure 2.9** Select direct targets of ABI3 associated with developmental processes in which ABI3 is involved. ***, stage dependent, please see text.

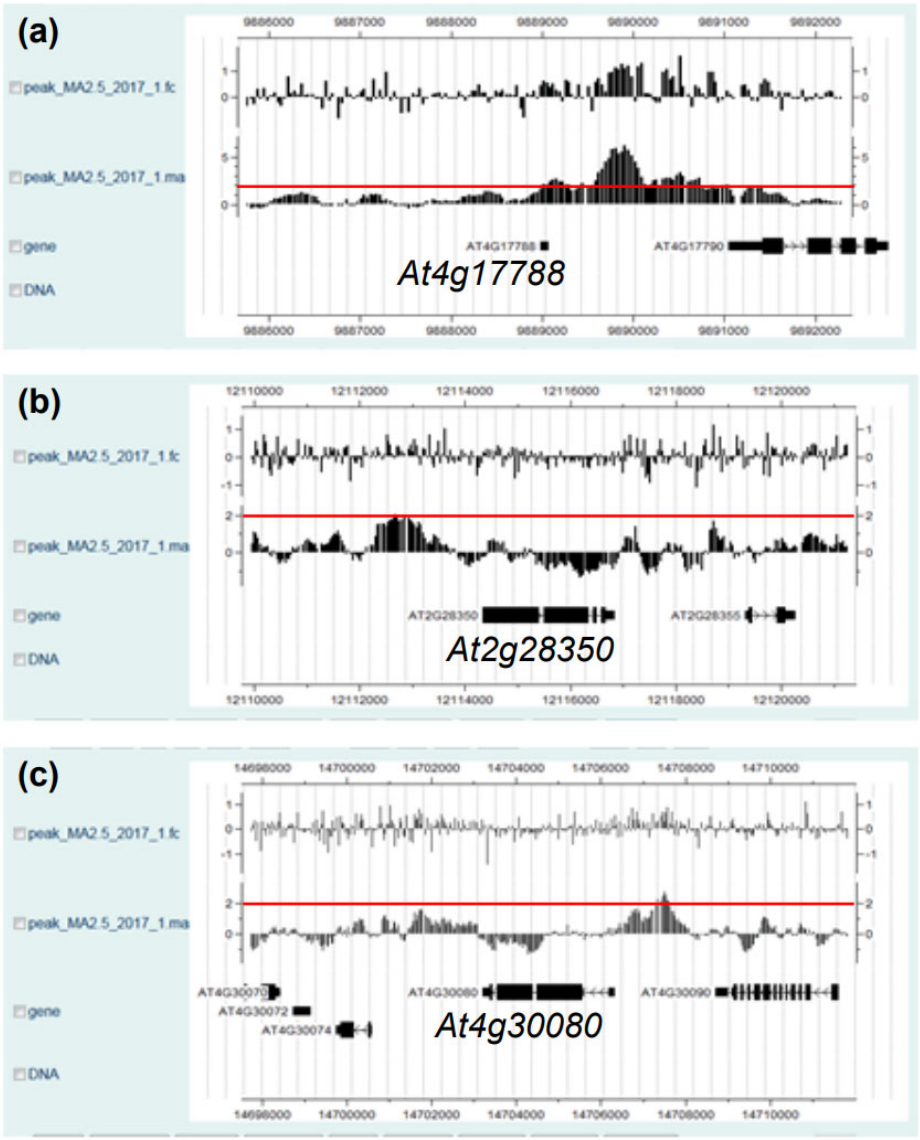



**Figure 2.10** Cisgenome traces showing potential ABI3 binding to regulatory regions associated with *At4g17788, miR160B* (a), *At2G28350, ARF10* (b), and *At4g30080, ARF16* (c). The red line indicates a 2-fold moving average comparing DNA recovered by immune ChIP to no antibody control.

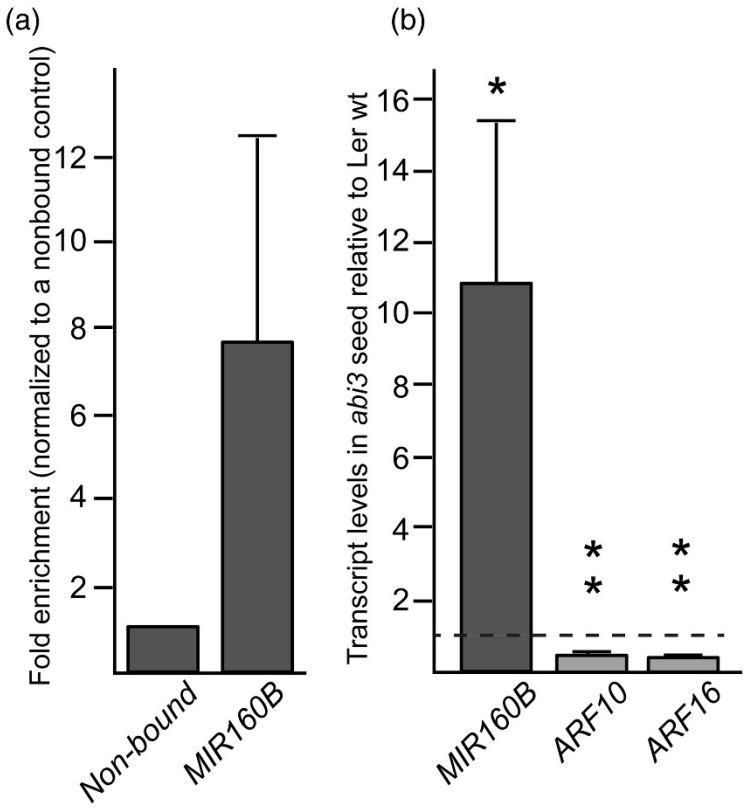



**Figure 2.11** The miRNA-encoding gene *MIR160B* is a directly repressed target of ABI3. (a) Fold enrichment of regulatory regions associated with *MIR160B* from ChIP using HA antibody and ABI3-HA tissue. Data is normalized to a non-bound region of the genome. (b) Transcript accumulation corresponding to *pri-miRNA160B, ARF10* and *ARF16* in *abi3-5* developing seeds relative to Ler wild type set to 1 (dotted line). Means and SE of the means are shown for three biological replicates. *significant at P<0. 05; ** at P<0.01

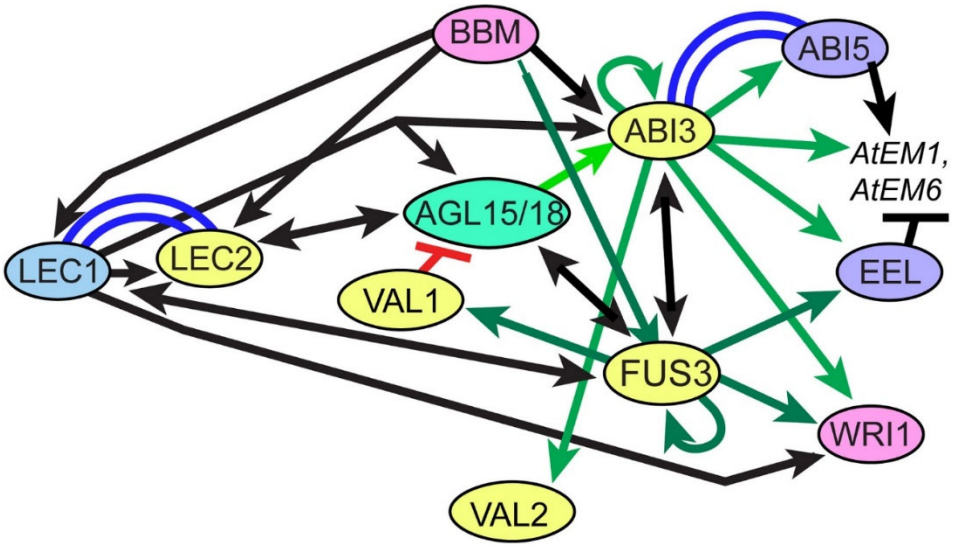



**Figure 2.12** Working model summarizing interactions between select key regulators of embryogenesis, somatic embryogenesis (SE) and the transition to seedling development. Solid lines with arrows at both ends indicate a positive regulatory loop. Other select directly expressed genes are shown by black or green arrows. ABI3 targets are from this study and the remainder are from other published data (LEC1, (Pelletier, Kwong et al. 2017); LEC2, (Braybrook, Stone et al. 2006); FUS3, (Wang and Perry 2013); VAL1/HSI2, (Chen, Veerappan et al. 2018); BBM, (Horstman, Li et al. 2017); AGL15, (Zheng, Ren et al. 2009); ABI3-ABI5 protein interaction diagrammed by double blue line, (Nakamura, Lynch et al. 2001); LEC1-LEC2 interaction, (Boulard, Thevenin et al. 2018); regulation of AtEM1 and 6 by ABI5 and EEL, (Bensmihen, Rippa et al. 2002). Yellow, B3 domain protein; purple, basic Leu zipper; pink, AP2 related; green, MADS; blue, NFYB/HEME-ACTIVATED PROTEIN3. Not all interactions or interesting proteins are diagrammed or relevant references cited in the interest of clarity.

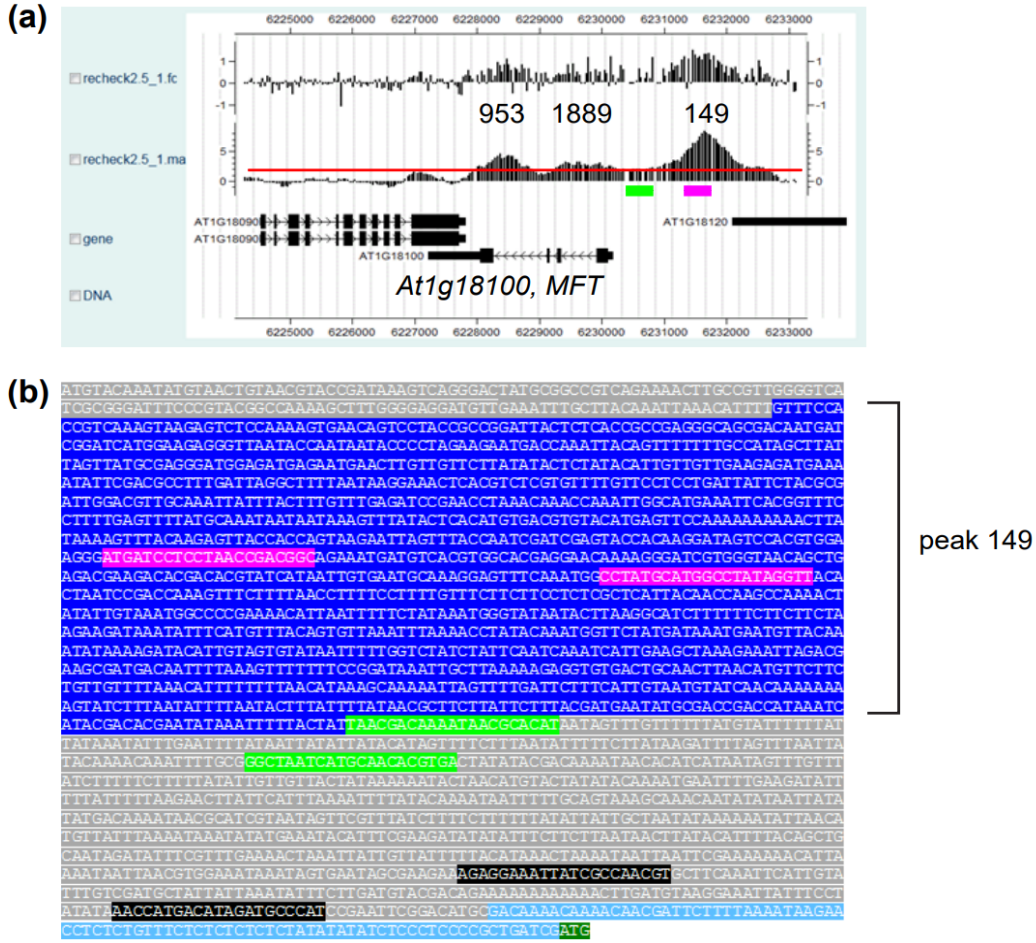



**Figure 2.13** Binding of ABI3 to *MFT* regulatory regions. (a), Cisgenome trace showing potential ABI3 binding to regulatory regions associated with *MFT (At1g18100)*. The red line indicates a 2-fold moving average comparing DNA recovered by immune ChIP to no antibody control. The numbers above the peaks indicate the rank of the peak as identified by Cisgenome. (b), Region upstream of the start codon (dark green highlight) of *MFT*, including the region bound by ABI3 (blue highlight, peak 149). The pink and light green indicate the primers used in Xi et al., 2010 to amplify MFT-2 and MFT-3 respectively. Approximate location of these regions is indicated by the pink and green lines on A (not to scale).

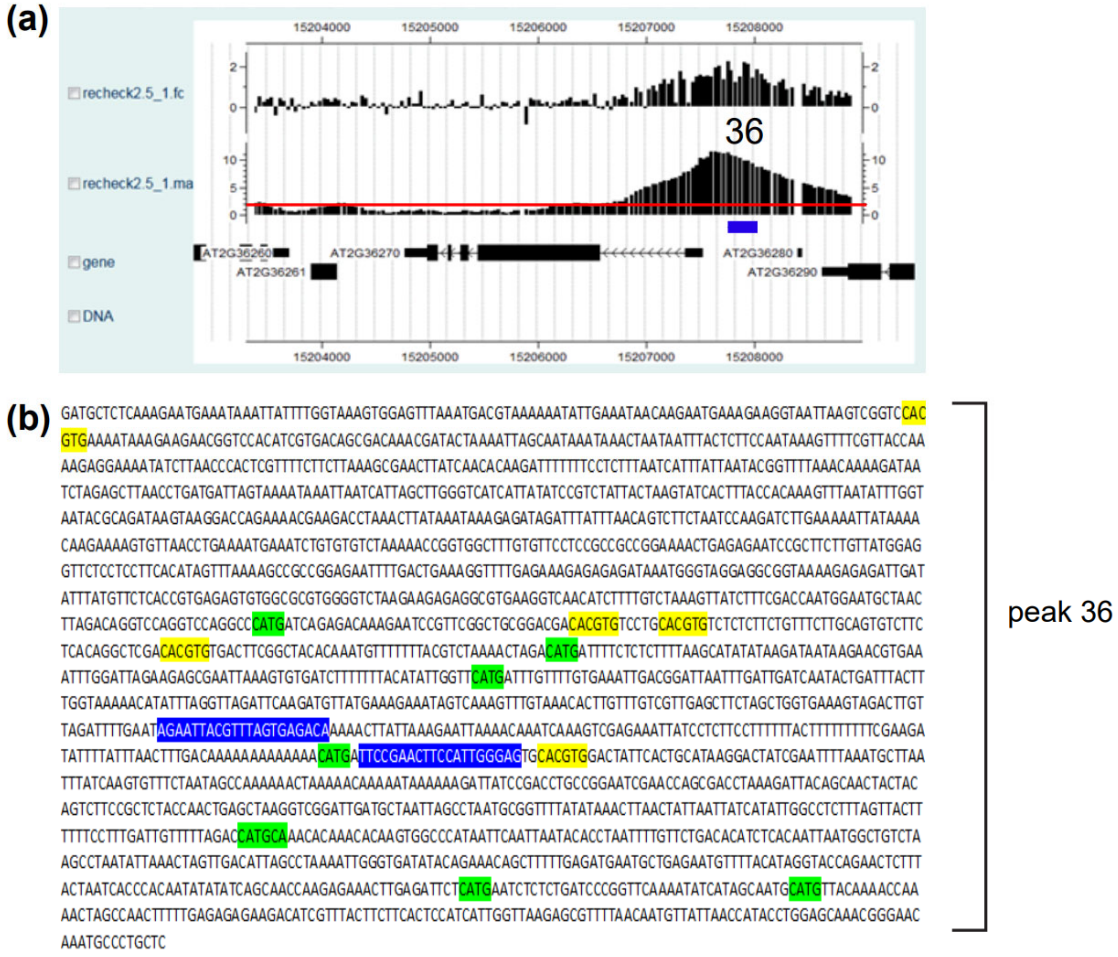



**Figure 2.14** Binding of ABI3 to *ABI5* regulatory regions. (a) Cisgenome trace showing potential ABI3 binding to regulatory regions associated with *ABI5 (At2g36270)*. The red line indicates a 2-fold moving average comparing DNA recovered by immune ChIP to no antibody control. The number above the peak indicates the rank of the peak as identified by Cisgenome. (b) Sequence of the regions bound by ABI3. The blue highlights indicate the primers used in Xi et al., 2010 to amplify ABI5-2. Approximate location of this region is indicated by the blue line on A (not to scale).

## Chapter 3 *Arabidopsis thaliana* SIN3A associated polypeptide P18 (AtSAP18) plays a role in response to ABA

### 3.1 Abstract

*Arabidopsis thaliana* SIN3A ASSOCIATED POLYPEPTIDE P18 (AtSAP18), an orthologue of human SAP18, is a transcriptional co-regulator that is a component of histone deacetylase (HDAC) complexes. Previous studies showed that expression of *AtSAP18* is induced by salt, drought, abscisic acid (ABA) and ethylene treatment, and that the protein encoded by this gene may be involved in the regulation of salt stress. Here I assessed transcriptomes in response to SAP18 by comparing *sap18* mutant and wild type (WT) developing seeds and seedlings by using high-throughput RNA sequencing (RNA-seq) experiments. Also, I report on chromatin immunoprecipitation sequencing (ChIP-seq) experiments to globally map genomic sites with which SAP18 associates. By combining these two large data sets, I could identify the genes that are potentially directly regulated by SAP18. Gene ontology (GO) term analysis of direct SAP18 targets is consistent with a role for SAP18 in gene regulation in response to several abiotic stress. Also, I found significant enrichment of genes involved in ABA response regulated by SAP18 in developing seeds and/or seedling tissue. More importantly, I discovered a new phenotype of a loss-of-function mutant *sap18* in that the mutant is hypersensitive to ABA treatment compared to Columbia (Col) (WT), suggesting an important role of SAP18 in modulation of ABA response. Interestingly, *DELAY OF GERMINATION 1* (*DOG1*) is up regulated 2.57 times in the *sap18* mutant compared to WT in seeds. On the contrary, *DOG1* is repressed 4.7 times in *sap18* compared to WT in the seedling samples. This may suggest a role for

SAP18 in control of developmental programs in the transition from seed to seedling development.

### 3.2 Introduction

Unlike animals' mobility that allows them to escape environmental pressures, plants must adjust their physiology to adapt to different environmental challenges including various abiotic and biotic stresses. Drought is one of the most extensive and harmful environmental stresses that plants have to handle and will have an increasingly profound influence on crop yield due to ongoing global climate change (Shanker, Maheswari et al. 2014). It is estimated that drought will result in serious plant growth problems for more than 50% of the arable lands by 2050 (Ashraf and Wu 1994, Vinocur and Altman 2005, Kasim, Osman et al. 2013, Vurukonda, Vardharajula et al. 2016), a problem made even more challenging by the need to feed a growing population.

Abscisic acid (ABA) is a plant hormone, which acts as a crucial messenger that merges environmental signals and internal metabolism response. ABA plays an essential role in controlling plants' metabolic response to environmental stresses such as salinity, UV, and drought. It is considered as a major phytohormone in response to drought through changing physiological and molecular processes including root development and stomatal closure. It also regulates various plant growth and development including seed dormancy, germination, leaf senescence, embryo, maturation, and floral induction (Tuteja and behavior 2007, Sah, Reddy et al. 2016, Vishwakarma, Upadhyay et al. 2017). Additionally, other plant hormones such as auxins (IAA), brassinosteroids (BRs), cytokinins (CKs), ethylene (ET), gibberellins (GAs), jasmonic acid (JA), and salicylic acid (SA) are also

important for plants to adapt to drought conditions (Finkelstein 2013, Sah, Reddy et al. 2016, Ullah, Manghwar et al. 2018).

Gene regulation in response to stress is mediated in part by epigenetic modification (Sokol, Kwiatkowska et al. 2007). Epigenetic modification of nucleosomes (a unit of chromatin) can have both positive and negative influences on gene transcription The dynamic effects of nucleosome function are greatly impacted by post-translational modification, such as phosphorylation, methylation, and acetylation. Acetylating and deacetylating highly conserved lysine residues on histones results in increased and decreased transcriptional activities, respectively. Lysine acetylation happens in all histones (H1, H2A, H2B, H3, and H4) and the positions are not random (Carmen, Rundlett et al. 1996, De Ruijter, Van Gennip et al. 2003). Here we are interested in AtSAP18 (for Arabidopsis Sin-Associated Protein of 18 kD, At2g45640) that is a component of histone deacetylase (HDAC) complexes. AtSAP18 is an orthologue of mammalian SAP18, which was previously reported to be involved in the regulation of salt stress in Arabidopsis and its expression is induced by not only salt, but also cold, drought, ET and ABA (Song and Galbraith 2006). Interestingly, AtSAP18 can interact with proteins that have Ethylene-responsive element binding factor-associated Amphiphilic Repression (EAR) motifs, which is the first reported active repression motif in plants (Kagale and Rozwadowski 2011). For example, a yeast 2-hybrid (Y2H) screen identified SAP18 can interact with AGAMOUS-like 15 (AGL15) which has a form of the necessary EAR motif (LxLxL; L is leucine and x is any other amino acid). Similarly, the interaction between AGL15 and SAP18 is confirmed by gel shift assay where addition of SAP18 results in a super shift of the AGL15-DNA probe

complex, although SAP18 cannot bind to DNA itself and does not interact with another MADS-domain protein, SEPALLATA3 (SEP3) bound to DNA (Hill, Wang et al. 2008). Also, SAP18 interacts with key histone deacetylases, HDA6 and HDA19. AGL15 can also weakly interact with HDA19 (Song and Galbraith 2006, Hill, Wang et al. 2008). The Y2H screen and Bimolecular Fluorescence Complementation (BiFC) experiments show that SAP18 could interact with ABI5 binding protein 2 (AFP2) which is involved in the ABA response pathway (Lynch, Erickson et al. 2017). But little is known about the role of SAP18 in gene regulation involved in stress response pathways.

To further study the role of AtSAP18 involved in plants' stress response, I performed several RNA high-throughput sequencing experiments (RNA-seq) to identify its potential targets in developing seeds and young seedlings since *AtSAP18* transcript accumulates in both tissues. Also, it is important to understand the regulation pattern of AtSAP18 in the transition process before and after the seeds completing germination. Chromatin immunoprecipitation sequencing experiments (ChIP-seq) were also done in seedlings and embryo tissues to identify its direct targets and provides us a large data set, a very useful tool, to compare with the gene lists of other key transcription factors that interact with SAP18. I found several abiotic stress categories including response to salt and drought were overrepresented in the GO term analysis and many genes that are involved in an ABA response were in the list. Interestingly, I found one of the most critical genes in the ABA pathway, *DOG1*, was influenced by SAP18 oppositely in the two samples. ABI5 binding protein 1 (AFP1) was also regulated by SAP18 in seedling tissues. Moreover,

I compared the ABA sensitivities of the *sap18* mutant and Arabidopsis WT; and found *sap18* plants are hypersensitive to ABA.

## 3.3 Results and Discussion

### 3.3.1 Generation of Culture Tissue for ChIP

Embryo culture tissue (ECT) has been used previously to characterize the entire genome binding sites of the three embryo transcription factors, namely AGAMOUS-Like 15 (AGL15, a MADS domain transcription factor), FUSCA3 and ABI3 (both B3 domain transcription factors) (Zheng, Ren et al. 2009, Wang and Perry 2013, Tian, Wang et al. 2020). Unlike many other tissues, Arabidopsis developing seeds are challenging to use in ChIP experiments for subsequent high throughput sequencing due to the limited amount of material as well as the accumulation of compounds such as polysaccharides and oils that impede good sample preparation (Haque, Han et al. 2018). ECT provides an abundant tissue source from which we can achieve robust ChIP and prior characterization shows embryo programs are expressed in ECT (Harding, Tang et al. 2003). ECT was initiated by culturing isolated developing zygotic embryos of *SAP18pro:SAP15-HA/c-MYC; 35Spro:AGL15* on MS medium. Within three weeks, secondary embryos are apparent on a fraction of the explants. These can be subcultured every two to three weeks, and due to the presence of the *35Spro:AGL15* transgene, will produce stable ECT. Importantly, overexpression of *AGL15* in developing seeds has no effects on expression of *SAP18* (as measured by accumulation of transcripts, unpublished high throughput RNA-seq data; specifically there was a -1.42 fold change in the overexpression compared to WT and this was not significant) so overaccumulation of SAP18 is not a concern. Moreover, the ECT

does not over accumulate AGL15 compared to zygotic embryos, despite the *35S* promoter (Wang, Tang et al. 2002).

### 3.3.2 Gene Expression in response to SAP18

After receiving *sap18* loss-of-function mutant seeds from GABI-KAT (https://www.gabi-kat.de/) the homozygote plants were selected. Although the T-DNA insertion is located in an intron (Figure 3.1), SAP18 transcript was totally absent according to semi quantitative RT PCR results (Figure 3.2), indicating the allele obtained was a knock-out. RNA was extracted from *sap18* and WT as described in Methods, then RNA was sent to a company for library preparation and RNA-Seq. The RNA-Seq (mRNA-sequencing) (Novogene) was used to assess the transcriptome in response to SAP18 accumulation in WT and *sap18*. Three biological replicates were performed for each genotype and the percentage (%) of total mapped with genes for each replicate was determined. I found mapped reads, that serves as a measure of sequencing accuracy and lack of contaminating DNA, fell into the expected range of 70-90% (Conesa, Madrigal et al., 2016) (Table 3.1). Then I compared developing seeds collected from 7 to 8 days after flowering (daf) siliques of *sap18* to those of Col wild type (WT) and found a total of 87 genes to be significantly different between these two groups. Seventy genes show increased transcripts in 7 to 8 daf developing seeds of *sap18* to those of Col WT and 17 genes show decreased transcript levels for this comparison by using p-value <0.05 and fold change at least 2-fold cutoffs. Comparing 7 to 8 day-old (already imbibed and moist chilled for 3 days) seedlings of *sap18* to those of WT using the same cutoffs, there were 163 genes

with altered transcript accumulations. There are a total 101 genes that are up-regulated, and 62 genes are down-regulated when comparing seedlings of *sap18* to those of WT.

Gene ontology (GO) term analysis was performed using PANTHERTM (Mi, Poudel et al. 2016) and overrepresented gene ontology categories from "Biological Processes" are shown in Figure 3.3a and b respectively for SAP18 regulated genes in developing seeds and seedlings. Overrepresentation of genes involved in "seed maturation" (23.32 FE) (Figure 3.3A) in developing seeds, may suggest an important role of SAP18 in embryo development. In seedling tissue, the overrepresented category "response to salt stress" (4.76 fold enrichment (FE) over the genome) (Figure 3.3B) is notable as it agrees with the previously described *sap18* plant sensitivity to salt (Song and Galbraith 2006), which was specifically that the mutant plants produced less chlorophyll compared with the wild type when NaCl was present in the medium.

Interestingly, overrepresentation of genes in GO categories "response to abscisic acid" (5.38 FE seeds, 5.01 FE seedlings), "response to alcohol" (5.35 FE seeds, 4.98 FE seedlings), "response to acid chemical"(3.64 FE seeds, 4.98 FE seedlings ), and "response to oxygen-containing component"(3.33 FE seeds, 3.22 FE seedlings) are present in both developing seeds and seedlings samples (Figure 3.3A compared to B). Previous study showed *SAP18* transcript level was induced by ABA in leaf, steam, flower and root tissue of *Arabidopsis* (Song and Galbraith 2006). So, I decided to test the sensitivity of mutant plants and WT to ABA based on this result.

### 3.3.3 SAP18 is involved in the regulation of ABA response

Because "response to ABA" was an overrepresented GO category for SAP18 regulated genes, I next tested whether *sap18* had any ABA-related phenotypes. Seeds of *sap18* and WT control were allowed to complete germination on MS medium with and without ABA and the development of green cotyledons scored as a measure of seedling establishment. Here I would like to clarify that there were no significance *sap18* compared with WT for seeds complete germination. As shown in Figure 3.4, the number of seedlings with green cotyledons was reduced in *sap18* compared to WT when ABA was added to the medium. When relatively lower concentration (300 nanomolar(nM) and 500 nM) of ABA were used, *sap18* mutant developed fewer green cotyledons compared with WT at the fourth day after three days moist chilling to break dormancy. Forty percent of the *sap18* seedlings developed green cotyledons that was significantly lower (p-value<0.05, t-test) than the 68% WT seeds with green cotyledons when 300 nM ABA is used. Also, I had similar results when 500 nM ABA was applied in that 18% *sap18* seedlings developed green cotyledons while 44% WT formed green cotyledons. At this time 90% of s*ap18* plants and 100% WT plants in the medium without ABA had already developed green cotyledons (Figure 3.4A). But given additional time, the *sap18* mutants would get closer to WT with similar number of seedlings having green cotyledons. Thus, I also tested the ABA sensitivity using higher ABA concentration (3000 nM). Under this condition, at the sixteen days after the cold treatment, we could observe an obvious significant difference between *sap18* and WT. Only 23% of *sap18* mutant seeds formed green cotyledons while 70% showed this development for the WT (Figure 3.4B). In short, *sap18* mutant is

hypersensitive to ABA for seedling establishment but not for completion of germination, which suggests this gene plays an important role in regulation of ABA response.

As mentioned above, the "response to ABA" category is over-represented in the GO-term analysis for both young developing seeds and seedlings samples. There are ten genes and eighteen genes of seed and seedling samples, respectively, belong to this ontology category. To understand the possibility that how SAP18 affects plant response to ABA, I have listed these genes and their expression patterns in my samples (Table 3.2). Initially confusingly, *SAP18* transcript is up regulated in the *sap18* mutant. Upon further inspection, I found that while there is reduced or no complete transcript (Figure 3.2), there were an increased number of RNA-seq reads at 3' of the T-DNA insertion position, possibly due to regulatory elements within the T-DNA.

Interestingly, I found three genes present in both tissue comparisons after removing *SAP18* itself from the list; they are *AT2G32510, MITOGEN-ACTIVATED PROTEIN KINASE KINASE KINASE 17 (MAPKKK17); AT5G45830, DELAY OF GERMINATION 1(DOG1)* and *AT2G47770, TSPO* (*OUTER MEMBRANE TRYPTOPHAN-RICH SENSORY PROTEIN)-LIKE PROTEIN*. DOG1 is one of the most crucial transcription factors that is involved in the seed ABA regulatory pathway. The *dog1* mutants are completely nondormant and shows a reduced seed longevity phenotype (Bentsink, Jowett et al. 2006). Here I found *DOG1* was up regulated 2.57 times in *sap18* mutant compared WT in the seeds. On the contrary, *DOG1* is repressed 4.7 times in *sap18* compared to WT in the seedling samples (Table 3.2).

### 3.3.4 Genome-Wide Identification of Indirect, in Vivo Binding Sites for SAP18

SAP18 is believed to function as a co-factor/regulatory protein that associates with other transcription factors to achieve its functions, and therefore does not bind directly to DNA but rather is indirectly associated via a DNA bound transcription factor. During chromatin immunoprecipitation (ChIP) assays, formaldehyde is used;, and as a result, not only protein and DNA is crosslinked but also proteins with proteins, which makes it possible to perform a ChIP-seq for SAP18. To locate the indirectly associated genes of SAP18, we use the ChIP-seq method, in which anti-c-MYC antibody was used to immunoprecipitate the SAP18-c-MYC-unknown proteins-DNA complex, and the recovered DNA is used for high-throughput sequencing (BGI group company). Western blot confirmed that the desired protein was present during the ChIP process and sufficiently precipitated and recovered at the last elution step from the beads(Figure 3.5). A clear band corresponding to SAP18-c-MYC in the final sample is present only when the anti-c-MYC antibody was used for ChIP, andnot in the no antibody control (Figure 3.5 last two lanes). CLC genomic benchwork (Epigenomic Analysis – ChIP-Seq Analysis) was used to analyze the ChIP-seq data and the workflow used was that described in the CLC Manual using default settings (https://resources.qiagenbioinformatics.com/tutorials/ChIP-seq_peakshape.pdf). The quality measures for each replicate was shown in Table 3.3. When I was matching the reads to the reference (TAIR10), the default setting parameters used were as follows: 1) Match core 1, 2) Mismatch cost 2, 3) Linear gap cost, 4) Insertion cost 3, 5) Deletion cost 3, 6) Length fraction 0.5, 7) Similarity fraction 0.8, and 8) Non-specific match handing Ignore thresh holds. I analyzed the results from the read mapping

to detect significant peaks with Maximum P-value for calling to the value of 0.05 instead of the default 0.1 since p-value less than 0.05 is a threshold that is more well accepted and strict. An example of the peak (annotated to AT5G45830 DOG1) that is associated with SAP18 is shown in Figure 3.6.

I did three biological replicates for each sample including ECT and seedling tissue. Transcription factors could bind not only to the estimated promoter region of a gene but also 3' prime end and within the gene (including introns). Thus, the peaks were counted in all the positions to avoid missing any binding regions. There are 3692, 6515 and 1324 genes respectively found in the ECT three replicates and 12647, 1627 and 6317 genes respectively discovered in the seedling replicates (Table 3.4). Even though the sequencing data has good quality, the replicates of the same tissue still have differences in the peaks. Some reports have indicated that two biological replicates for a ChIP experiment are sufficient, but some important binding sites are likely to be lost if only two sets of duplicates are done. Adding the third biological repetition definitely increases reliability of peak identification. I followed a simple rule that if the majority of samples (in our case is two out of three) identify a peak, then this peak is a potential real target (Table 3.4 Majority rule line). Unlike the strategy that a true peak must be identified in all replicates, this simple rule could avoid some important peak positions that failed to be detected due to one replicate that has low reads or high noise background (Yang, Fear et al. 2014). In short, we found 3047 genes in the ECT and 5045 genes in the seedling material. Similar with the RNA-seq results, there are more targets in the seedlings compared with embryo type tissue (Table 3.4). Then I combined the RNA-seq and ChIP-seq data to find the targets

that were not only regulated but also associated with SAP18. Twenty-nine common genes are found in the seedlings and 14 common genes are found in the embryo tissue (seeds for RNA-seq and ECT for ChIP-seq). Among them, there are two genes in all the lists and only one left if we removed SAP18 itself (Figure 3.7). It is *AT3G30720*, *QUA-QUINE STARCH* (*QQS*).

**3.4 Conclusion and Future work**

I report on the targets regulated by SAP18, revealing that SAP18 functions in abiotic stress. We discovered a new *sap18* mutant phenotype that indicates SAP18 is involved in the regulation of ABA response. Also, we found a series of candidate genes related to the ABA response pathway that were regulated by SAP18. We confirmed that performing ChIP experiments on a co-regulator is possible in different tissues and a group of candidate genes are found to be directly responsive to SAP18 by combining the high through-put RNA-Seq and ChIP-Seq data. In this study, AT1G69260/ABI5 binding protein 1 (AFP1) is found to be up regulated in the *sap18* mutant of 7 to 8 day seedlings. This is inconsistent with a previous study that showed *afp1* mutant is strongly hypersensitive to ABA and over-expression of *AFP1* is insensitive to ABA due to its negative regulation (facilitating ubiquitin-mediated proteolysis) to ABI5 protein level (Lopez-Molina, Mongrand et al. 2003). Interestingly only AFP2 was reported to interact with SAP18 based on Bimolecular Fluorescence Complementation (BiFC) despite the fact that all members of the AFP family have the EAR motif (identified by the sequencing of either LxLxL or DLNxxP) that is believed as an important trait of the proteins which could interact with SAP18. However, other proteins, such as TOPLESS (TPL) and TOPLESS RELATED PROTEINS

(TPRs) also interact with EAR motifs. SAP18 and TPL/TPRs then recruit histone deacetylases such as HDA19 and HDA6 to repress the target genes (Kagale and Rozwadowski 2011). Possibly, some proteins only interact with one or the other classes of these proteins. At the same time, TPL and HDA19 could interact with both AFP1 and AFP2 (Lynch, Erickson et al. 2017) and AGL15 was also found to interact with TPL and one of the TPRs. Thus, it is necessary to use another method, for example the more reliable CoIP, to confirm these interactions and also to find novel connections among these proteins in the future work. Also, other candidate genes related to ABA in Table 3.2 need be further studied, for example the PYR1-LIKE4 (PYL4) is significantly down-regulated in the *sap18* mutant compared with WT in this study. PYL4 has a role in regulating the ABA signaling pathway (Lackman, González-Guzmán et al. 2011, Pizzio, Rodriguez et al. 2013). *DOG1* is a key transcription factor that involved in ABA response pathway, 100% seeds will complete germination at the first day without *DOG1* after moist chilling (Bentsink, Jowett et al. 2006). DOG1 can promote *ABI5* activity thus further activates transcription of many *LEA* and *HSP* genes (Skubacz, Daszkowska-Golec et al. 2016). Here, the *DOG1* transcript is regulated oppositely by SAP18 in developing seeds and seedlings, thus it is important to test when this switch occurs in the future work.

## 3.5 Methods

### 3.5.1 Transgene Construct

For the *SAP18pro:SAP18-c-MYC or HA* constructs, about 1 kb 5' of the start codon of SAP18 and the entire genomic region excluding the stop codon (~2.4 kb total) was amplified and cloned into pENTR/D-TOPO vector (Invitrogen) following the

manufacturer's instructions (SAP18 native promoter 5'-CACCAGTCGGCGTGTCAGTGTCTTGTGG-3'/end of the gene without stop code: 5'-GTAAATTGCCACATCCAGATAATC-3'). The insert was moved into the destination vectors pGWB19 (10x c-MYC) or pGWB13 (3x HA) (Nakagawa, Ishiguro et al. 2009); obtained from Dr. T. Nakagawa, Shimane University) following the manufacturer's instruction for Gateway LR Clonase II Enzyme mix (Invitrogen). More details for generating materials for ChIP are described in the Results.

### 3.5.2 Plant Material

*Arabidopsis* Col WT, and *sap18* plants were grown as described in (Wang and Perry 2013). To establish embryonic culture tissue (ECT) for ChIP experiments, a *35Spro:AGL15* transgene was introduced by crossing with the *SAP18pro:SAP18-c-MYC* lines. Developing zygotic embryos of green bent cotyledon stage were isolated and placed into culture as described in Harding et al. (2003).

### 3.5.3 RNA Sequencing

Flowers of Col WT and *sap18* were tagged on the day that they opened. Seeds were collected at 7-8 day after flowering (daf), and flash frozen in liquid nitrogen. For another type of tissue, seedlings were collected after addition of 7-8 days from the media after 4°C cold treated for 3 days, and flash frozen in liquid nitrogen. RNA was extracted and cleaned with the QIAGEN RNeasy Plant Mini Kit. Three biological replicates were prepared and sent to Novogene (California) for library preparation and RNA-seq. Data was analyzed by using CLC Genomics benchwork.

### 3.5.4 ChIP sequencing

ChIP was performed with the ? plant tissue as in (Wang and Perry 2013). The control was the same tissue, but ChIP performed without primary antibody. Three biological replicates were prepared and sent to the company BGI group (Hong Kong) for quality check, library preparation and ChIP-seq. Data was analyzed by using CLC Genomics benchwork, more details are described in the result part.

### 3.5.5 Genotyping

The *sap18* mutant used in this study is T-DNA insertion site (GABI_C85502). Primers 2341, 2241 and 1098 were used for genotyping. Primers 1097 and 2168 were used for RT-PCR. Primers 808 and 809 were positive control tubulin2 for RT-PCR. (2341 5'-AATGCTAGATTGTCTTTTGCGTTT-3'/2241 5'-ATAATAACGCTGCGGACATCTACATTTT-3'/1097 5'-AATGGTCTCCCACCACCTTGTCTT-3'/1098 5'-CTCAAACGGAAGTTCGGAAAGCGT-3'/2168 5'-GTAAATTGCCACATCCAGATAATC-3'/808 5'-GTCCTACTTTGTGGAGTGGA-3'/809 5'-CTGTGTACCAATGCAAGAA-3')

**Table 3.1** Fragment statistics for RNA-seq. The last column shows the percentage (%) total mapped to Arabidopsis genes. Uniquely mapped: A reads is uniquely mapped to one gene.

| Sample | Type | Uniquely mapped | Non-specifically mapped | Mapped | % of total mapped with type |
|---|---|---|---|---|---|
| Seeds WT population 1 | Gene (total) | 14248887 | 450060 | 14698947 | 93.08 |
| | Intergenic | 552479 | 540623 | 1093102 | 6.92 |
| | Total | 14801366 | 990683 | 15792049 | 100 |
| Seeds WT population 2 | Gene (total) | 17235658 | 209457 | 17445115 | 96.7 |
| | Intergenic | 483258 | 112555 | 595813 | 3.3 |
| | Total | 17718916 | 322012 | 18040928 | 100 |
| Seeds WT population 3 | Gene (total) | 13043945 | 179518 | 13223463 | 95.56 |
| | Intergenic | 480015 | 134916 | 614931 | 4.44 |
| | Total | 13523960 | 314434 | 13838394 | 100 |
| Seeds *sap18* population 1 | Gene (total) | 15918766 | 222534 | 16141300 | 97.44 |
| | Intergenic | 281179 | 143580 | 424759 | 2.56 |
| | Total | 16199945 | 366114 | 16566059 | 100 |
| Seeds *sap18* population 2 | Gene (total) | 17243538 | 255719 | 17499257 | 96.05 |
| | Intergenic | 570740 | 149522 | 720262 | 3.95 |
| | Total | 17814278 | 405241 | 18219519 | 100 |
| Seeds *sap18* population 3 | Gene (total) | 17828392 | 267862 | 18096254 | 96.21 |
| | Intergenic | 549993 | 162137 | 712130 | 3.79 |
| | Total | 18378385 | 429999 | 18808384 | 100 |
| Seedlings WT population 1 | Gene (total) | 12340805 | 385551 | 12726356 | 86.97 |
| | Intergenic | 690261 | 1216568 | 1906829 | 13.03 |
| | Total | 13031066 | 1602119 | 14633185 | 100 |
| Seedlings WT population 2 | Gene (total) | 13777723 | 826076 | 14603799 | 79.11 |
| | Intergenic | 1754904 | 2101667 | 3856571 | 20.89 |
| | Total | 15532627 | 2927743 | 18460370 | 100 |

**Table 3.1** (continued)

| | | | | | |
|---|---|---|---|---|---|
| Seedlings WT population3 | Gene (total) | 14290637 | 154958 | 14445595 | 90.81 |
| | Intergenic | 1150794 | 311102 | 1461896 | 9.19 |
| | Total | 15441431 | 466060 | 15907491 | 100 |
| Seedlings *sap18* population 1 | Gene (total) | 14182839 | 366355 | 14549194 | 89.55 |
| | Intergenic | 609110 | 1089253 | 1698363 | 10.45 |
| | Total | 14791949 | 1455608 | 16247557 | 100 |
| Seedlings *sap18* population 2 | Gene (total) | 15496582 | 178344 | 15674926 | 92.65 |
| | Intergenic | 910586 | 332316 | 1242902 | 7.35 |
| | Total | 16407168 | 510660 | 16917828 | 100 |
| Seedlings *sap18* population 3 | Gene (total) | 13970536 | 181046 | 14151582 | 95.08 |
| | Intergenic | 351311 | 381366 | 732677 | 4.92 |
| | Total | 14321847 | 562412 | 14884259 | 100 |

**Table 3.2** Genes that are related to ABA and regulated by SAP18 in 7 to 8 daf developing seeds or 7 to 8 days seedlings. NS means not significant. Using fold change>2, p-value<0.05 cutoff.

| AGI | TAIR 10 description | *sap18*/Col WT | |
|---|---|---|---|
| | | Seeds | Seedlings |
| AT5G45830 | DELAY OF GERMINATION 1 (DOG1) | 2.57 | -4.70 |
| AT2G32510 | MITOGEN-ACTIVATED PROTEIN KINASE KINASE KINASE 17 (MAPKKK17) | 2.08 | 3.19 |
| AT4G28520 | CRUCIFERIN 3 (CRU3) | 4.79 | NS |
| AT1G03880 | CRUCIFERIN 2 (CRU2) | 4.49 | NS |
| AT1G56600 | GALACTINOL SYNTHASE 2(GOLS2) | NS | 3.43 |
| AT1G05100 | MITOGEN-ACTIVATED PROTEIN KINASE KINASE KINASE 18(MAPKKK18) | NS | 3.58 |
| AT5G44120 | RMLC-LIKE CUPINS SUPERFAMILY PROTEIN(CRA1) | 3.43 | NS |
| AT4G21410 | CYSTEINE-RICH RLK (RECEPTOR-LIKE PROTEIN KINASE) 29(CRK29) | NS | -2.75 |
| AT2G45640 | SIN3 ASSOCIATED POLYPEPTIDE P18(SAP18) | 5.56 | 27.17 |
| AT2G39800 | DELTA1-PYRROLINE-5-CARBOXYLATE SYNTHASE 1(P5CS1) | NS | 2.35 |
| AT2G46680 | HOMEOBOX 7(HB-7) | NS | 2.23 |
| AT2G47770 | TSPO(OUTER MEMBRANE TRYPTOPHAN-RICH SENSORY PROTEIN)-LIKE PROTEIN(TSPO) | 3.31 | 4.04 |
| AT3G57540 | REMORIN FAMILY PROTEIN(AT3G57540) | NS | 2.69 |
| AT2G21490 | DEHYDRIN LEA(LEA) | 3.33 | NS |
| AT5G59220 | PP2C PROTEIN (CLADE A PROTEIN PHOSPHATASES TYPE 2C)(HAI1) | NS | 3.20 |
| AT2G38310 | PYR1-LIKE 4(PYL4) | NS | -2.32 |
| AT4G05100 | MYB DOMAIN PROTEIN 74(MYB74) | NS | 3.72 |
| AT5G50700 | HYDROXYSTEROID DEHYDROGENASE 1(HSD1) | 2.68 | NS |
| AT2G33380 | CALEOSIN-RELATED FAMILY PROTEIN(RD20) | NS | 3.20 |
| AT1G69260 | ABI FIVE BINDING PROTEIN(AFP1) | NS | 2.19 |
| AT5G14920 | GIBBERELLIN-REGULATED FAMILY PROTEIN(AT5G14920) | NS | -3.24 |
| AT4G21440 | MYB-LIKE 102(MYB102) | NS | 3.28 |

**Table 3.2** (continued)

| | | | |
|---|---|---|---|
| AT1G47510 | INOSITOL POLYPHOSPHATE 5-PHOSPHATASE 11(5PTASE11) | NS | 5.53 |
| AT5G52300 | CAP160 PROTEIN(LTI65) | 2.20 | NS |

**Table 3.3** Quality measures for ChIP-seq.

The relative strand correlation describes the ratio between the fragment-length peak and the read-length peak in the cross-correlation plot. This value should be greater than 0.8 for transcription factor binding sites but can be lower for ChIP-seq input or for histone marks.

The normalized strand coefficient describes the ratio between the fragment-length peak and the background cross-correlation values. This value should be greater than 1.05 for ChIP-seq experiments.

| Sample | Measure | Value | Statu |
|---|---|---|---|
| ECT with antibody population 1 | Number of reads | 9824092 | OK |
| | Relative strand correlation | 0.662 | Low |
| | Normalized strand coefficient | 1.105 | OK |
| ECT without antibody population 1 | Number of reads | 12084545 | OK |
| | Relative strand correlation | 0.39 | Low |
| | Normalized strand coefficient | 1.036 | Low |
| ECT with antibody population 2 | Number of reads | 8878000 | OK |
| | Relative strand correlation | 1.035 | OK |
| | Normalized strand coefficient | 1.547 | OK |
| ECT without antibody population 2 | Number of reads | 8261903 | OK |
| | Relative strand correlation | 0.98 | OK |
| | Normalized strand coefficient | 1.353 | OK |
| ECT with antibody population 3 | Number of reads | 18462302 | OK |
| | Relative strand correlation | 0.849 | OK |
| | Normalized strand coefficient | 1.102 | OK |
| ECT without antibody population 3 | Number of reads | 8408651 | OK |
| | Relative strand correlation | 1.074 | OK |
| | Normalized strand coefficient | 1.373 | OK |

**Table 3.3** (continued)

| | | | |
|---|---|---|---|
| Seedlings with antibody population 1 | Number of reads | 15966171 | OK |
| | Relative strand correlation | 1.022 | OK |
| | Normalized strand coefficient | 1.238 | OK |
| Seedlings without antibody population 1 | Number of reads | 14484746 | OK |
| | Relative strand correlation | 0.97 | OK |
| | Normalized strand coefficient | 1.112 | OK |
| Seedlings with antibody population 2 | Number of reads | 15215538 | OK |
| | Relative strand correlation | 0.913 | OK |
| | Normalized strand coefficient | 1.128 | OK |
| Seedlings without antibody population 2 | Number of reads | 9959274 | OK |
| | Relative strand correlation | 0.934 | OK |
| | Normalized strand coefficient | 1.152 | OK |
| Seedlings with antibody population 3 | Number of reads | 11438022 | OK |
| | Relative strand correlation | 0.799 | Lo |
| | Normalized strand coefficient | 1.151 | OK |
| Seedlings without antibody population 3 | Number of reads | 16485590 | OK |
| | Relative strand correlation | 0.956 | OK |
| | Normalized strand coefficient | 1.166 | OK |

**Table 3.4** Annotated gene numbers in the ChIP-seq results for SAP18 in ECT and seedling. Majority Rule: if the majority of samples (in our case is two out of three) identify a peak, then this peak is a potential real target.

| TISSUE TYPE | REPLICATE | GENE NUMBER |
|---|---|---|
| **ECT** | 1 | 3692 |
| | 2 | 6515 |
| | 3 | 1324 |
| | Majority Rule | **3047** |
| **SEEDLING** | 1 | 12647 |
| | 2 | 1627 |
| | 3 | 6317 |
| | Majority Rule | **5045** |

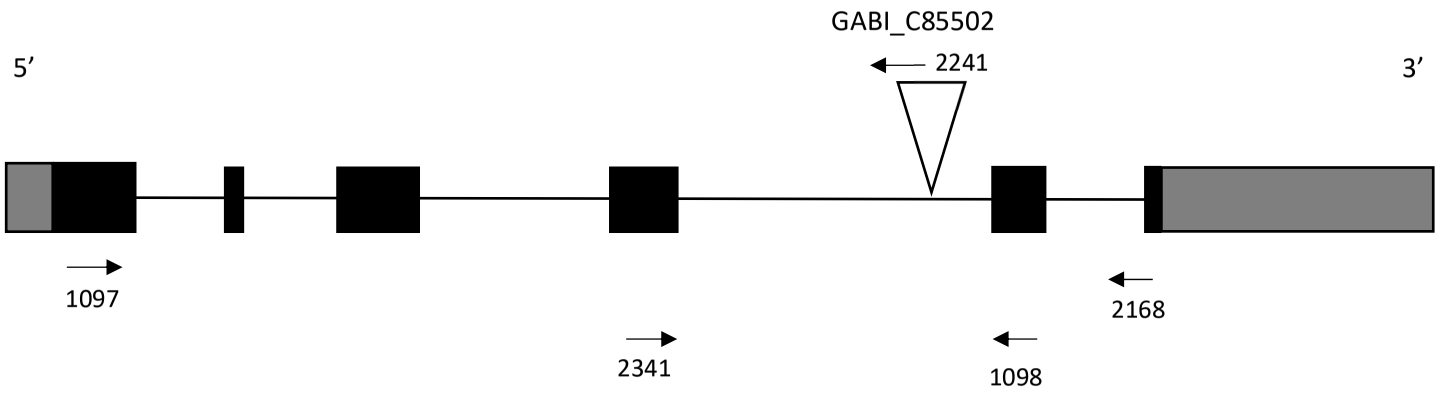





**Figure 3.1** The *SAP18* Gene Structure. Solid boxes stand for transcribed regions including protein coding (black) and untranslated region (gray). The T-DNA insertion site (GABI_C85502) for the mutant used in this study is shown. Primers 2341, 2241 and 1098 were used for genotyping. Primers 1097 and 2168 were used for RT-PCR.

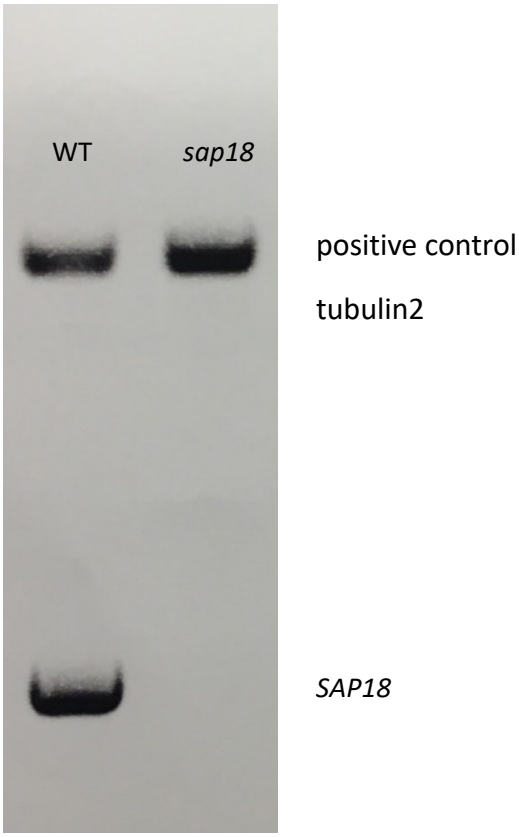



**Figure 3.2** DNA electrophoresis of RT-PCR samples show *SAP18* transcript was not detectable in the *sap18* mutant. This is one example from developing seeds. The first column is the Col (WT) and the second column is *sap18* mutant. The first row is positive control *TUBULIN2* that showed the RT-PCR works in both samples. The second row indicates the *SAP18* band is only present in the WT. Primers used were described in the method part.

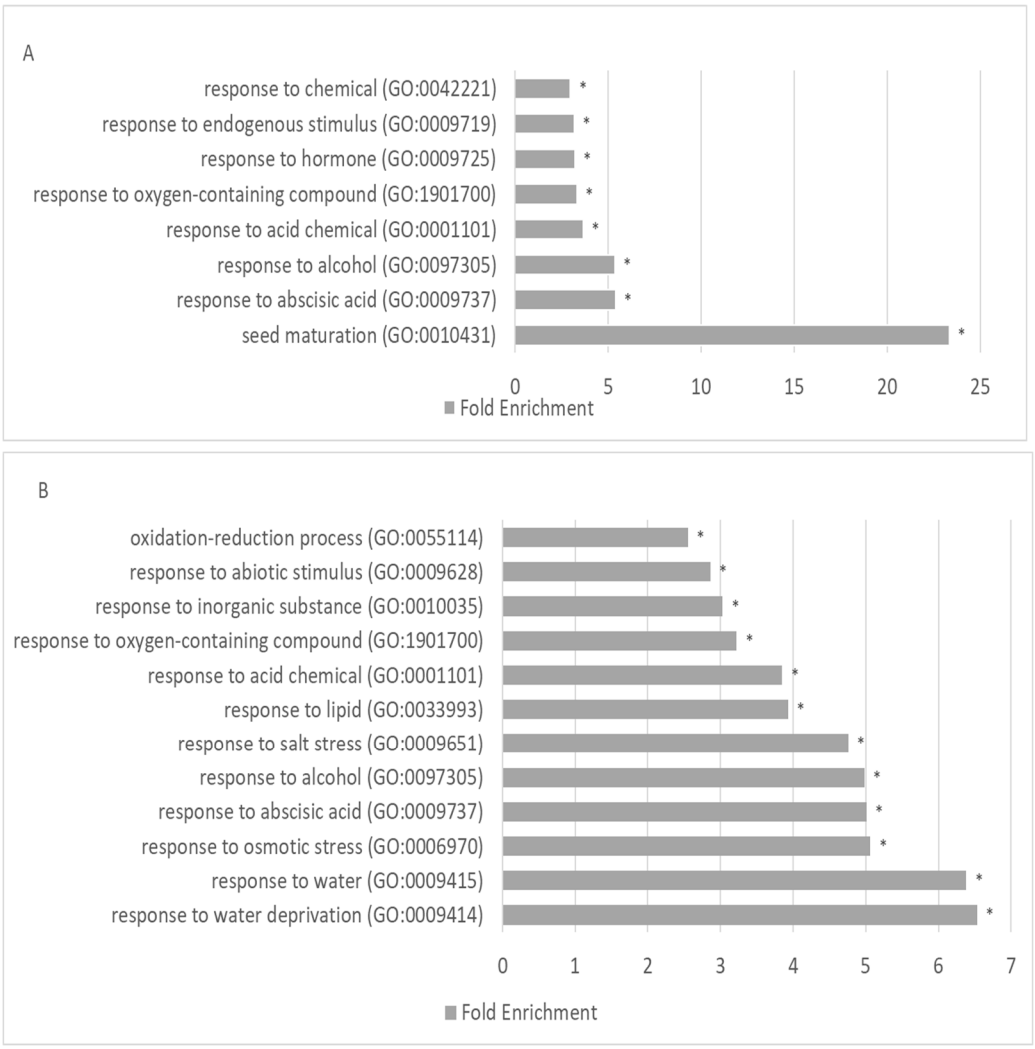



**Figure 3.3** Functional categorization of gene expression in response to *SAP18* using Panther classification system statistical overrepresentation test with default settings. All significant GO biological processes are shown as fold enrichment over the Arabidopsis genome.

(A) Response to SAP18 in 7 to 8 days after flowering (daf) developing seeds.

(B) Response to SAP18 in 7 to 8 days seedlings. * FDR<0.05

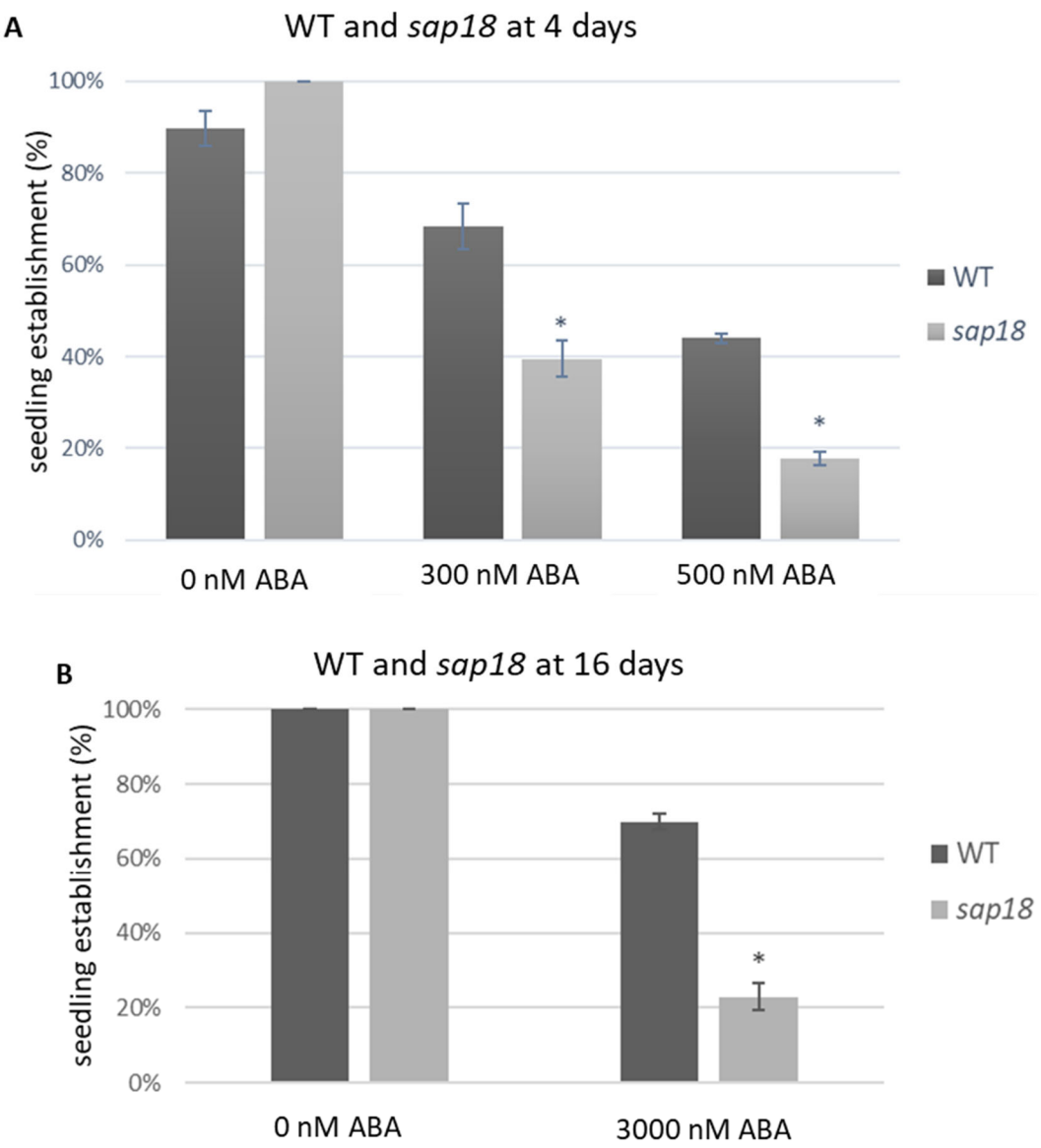



**Figure 3.4** Seedling establishment of the *sap18* and WT in response to ABA. Seeds of *sap18* and WT were sown on MS medium supplemented with different concentrations of ABA (A: 0 nM, 300 nM, 500 nM. B: 0 nM, 3000 nM) and seeds moist chilled for 3 days. Green cotyledon percentages were recorded at 4 days (A) and 16 days (B). Each value is the mean of at least 50 seeds from three independent experiments. Seeds of both genotypes plated on the same plate. Student's t-test was used to determine the significance between Col WT and *sap18,* * means p-value < 0.05.

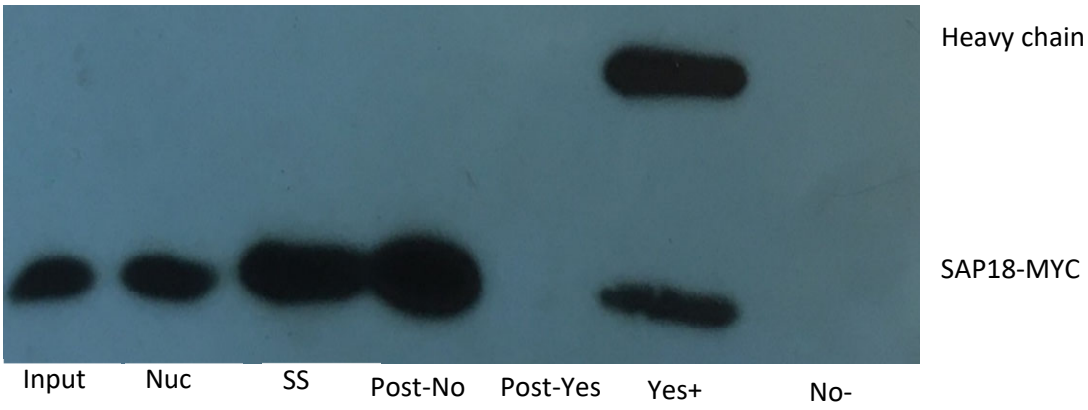



**Figure 3.5** Chemiluminescent Western blot using commercial anti-MYC antibody shows SAP18-MYC was precipitated sufficiently with the anti-MYC antibody. SAP18-c-MYC band size is around 35 to 42 kD (lower band). The high position band size is around 55kD and corresponds to the heavy chain of the anti-c-MYC antibody. Input, total protein; Nuc, nuclei; SS, sonication supernatant; Post-Yes, the SS sample with anti-SAP18 antibody from the supernatant after immunoprecipitation using protein-A Sephaose beads; Post-No, the SS sample without the anti-MYC antibody from the supernatant after mock immunoprecipitation; Yes+, Eluted protein from the beads when immunoprecipitation was done using MYC antibody; No-, negative control, eluted protein without the antibody. Heavy chain: non-specific reaction showed the heavy chain from the antibody itself. 1-minute exposure time. Western blot was performed as in (Wang, Tang et al. 2002).

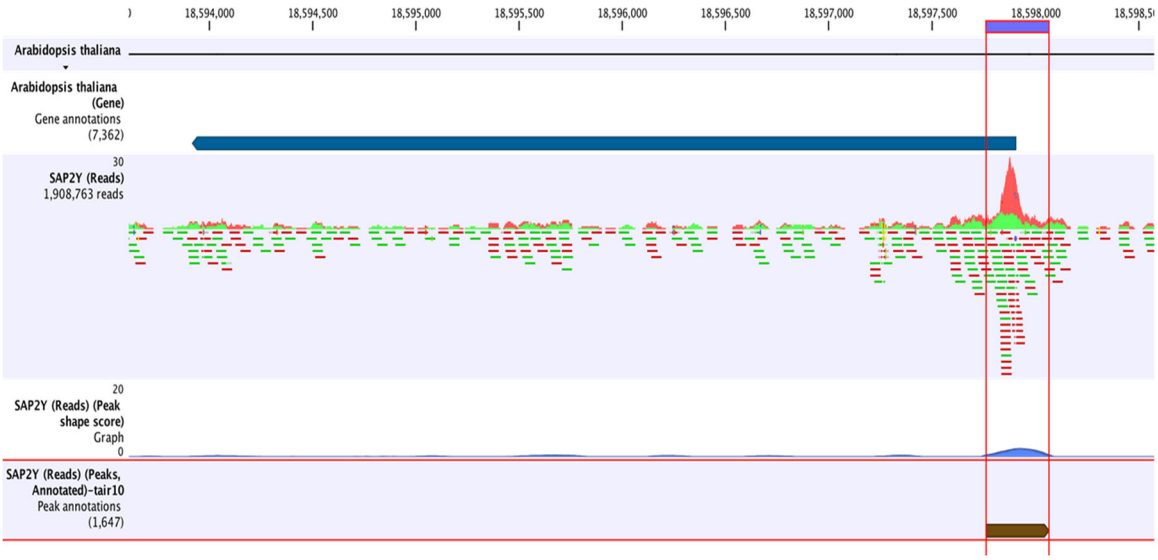
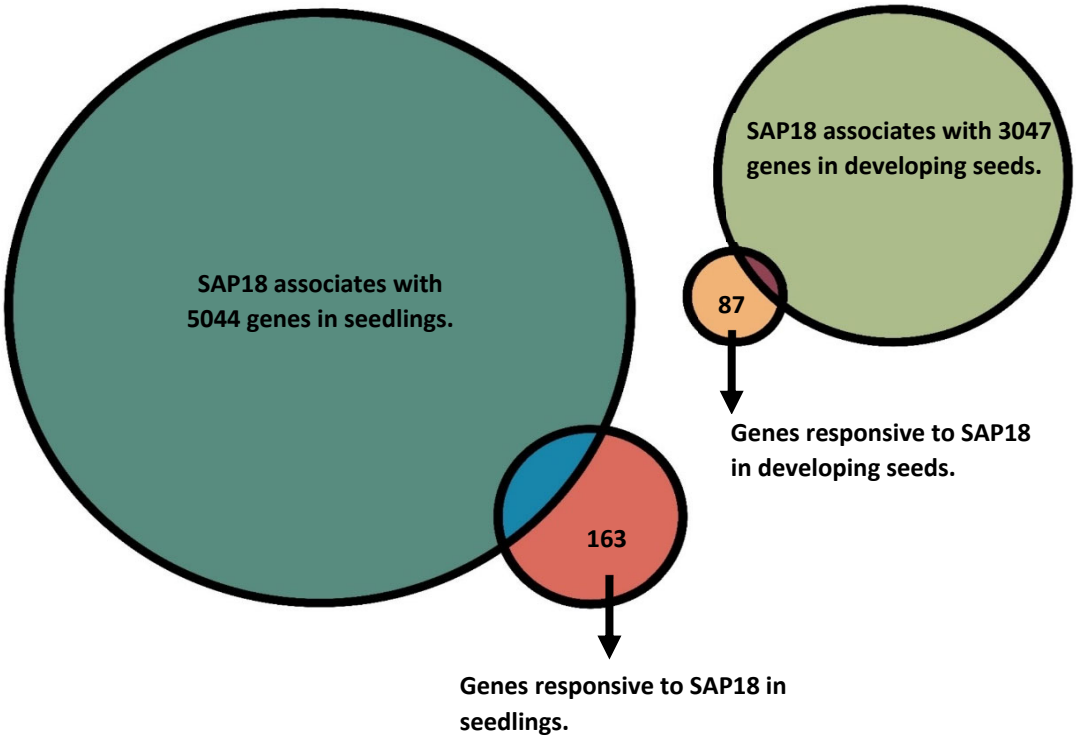



**Figure 3.6** ChIP-seq data, showed that the regulator SAP18 may associate with the regulatory regions of *AT5G45830 (DOG1)* in one population of ECT. The brown bar arrow is the annotated gene, short green and red lines are the reads from sequencing. Peak shape (blue) score is above 2, and p-value is below 0.05 as calculated by the CLC benchmark software described in (Strino and Lappe 2016).



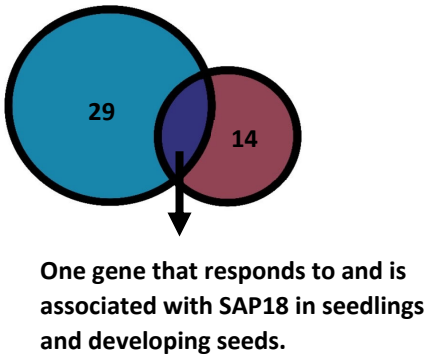




**Figure 3.7 Genes associated with SAP18.**

Blue: unique genes indirectly associated and regulated by SAP18 in seedlings. Brown: genes indirectly associated and regulated by SAP18 in developing seeds.

# Chapter 4 DUF1264 family plays a role in somatic embryogenesis and ABA response

## 4.1 Abstract

Many genes encode proteins with currently unknown function. In Arabidopsis, families of proteins that share conserved domains are designated as DUF (Domain of Unknown Function) and the family given a numerical identifier. The lab has been identifying genes directly regulated by key Arabidopsis embryo transcription factors AGL15, FUS3 and ABI3. Interestingly, all five members of the DUF1264 family may be directly expressed by different combinations of these transcription factors. DUF1264 is a seed specific protein family in Arabidopsis that their shared domain is highly conserved in various plants, such as tobacco, soybean, and rice suggesting that it plays an important role. Direct association of the transcription factors with these genes and assessing the effect on transcript accumulation by chromatin immunoprecipitation (ChIP)-qPCR and reverse transcriptase (RT)-qPCR, respectively, were performed in this study. A *DUF1264* quadruple mutant was created and showed a reduction in capacity of somatic embryogenesis (SE) compared to wild type control. Preliminary data showed an increase in SE for *DUF1264A* overexpression compared to wild type control. In addition, loss-of-function seedlings proved to be hypersensitive to ABA. A CRISPR-Cas9 design was successfully performed to generate the d*uf1264aabxccddee* pentuple mutant.

## 4.2 Introduction

The lab's approach to investigating embryogenesis involves identifying gene networks regulated by embryo transcription factors. Specifically, the lab has

performed chromatin immunoprecipitation (ChIP)-chip and expression microarray experiments to determine direct and indirect targets (respectively) of AGAMOUS-Like 15 (AGL15), FUSCA3 (FUS3) and ABSCISIC ACID INSENSITIVE 3 (ABI3) (Zheng, Ren et al. 2009, Wang and Perry 2013, Tian, Wang et al. 2020). AGL15 is a member of MADS-domain family of regulatory factors that is preferentially expressed during embryogenesis in *Arabidopsis thaliana* (Heck, Perry et al. 1995). It is also expressed in shoot apical meristems, young floral buds and leaf primordia, and roles in abscission, senescence and control of the transition to flowering have been reported (Fernandez, Heck et al. 2000, Adamczyk, Lehti-Shiu et al. 2007, Patharkar and Walker 2015). AGL15 accumulates in the developing Arabidopsis embryo and can directly bind regulatory regions of many targets some of which are relevant for embryogenesis including *ABI3* and *FUS3* (Wang, Caruso et al. 2004, Zheng, Ren et al. 2009). FUS3 is a B3 domain transcription factor that is a member of the LEAFY COTYLEDON (LEC) group of genes (Luerssen, Kirik et al. 1998). It is a regulator of gene expression during late embryogenesis (Keith, Kraml et al. 1994). ABI3 is a B3 domain transcription factor that is highly conserved, plays an important roles in seed maturation and it may control embryo degreening (Delmas, Sankaranarayanan et al. 2013).

Previous microarray data suggests that DUF1264 family members (genes) may be directly regulated by AGL15, ABI3 and/or FUS3 (Zheng, Ren et al. 2009, Yamamoto, Kagaya et al. 2010, Wang and Perry 2013, Tian, Wang et al. 2020). DUF1264 is a protein family of unknown function in which all five family members have the same conserved protein domain (duf, domain of unknown function number 1264). The expression

pattern is seed specific as measured by transcripts accumulation for all members (Klepikova, Kasianov et al. 2016). The five members in Arabidopsis are designated here as DUF1264A(AT1G05510), DUF1264B(AT2G31985), DUF1264C(AT4G18920), DUF1264D(AT5G45690), and DUF1264E(AT1G29680). DUF1264 is highly conserved in various plants, such as Arabidopsis, tobacco, soybean, and rice, which suggests that it plays an important role in plants. Interactions (protein- promoter [direct column in the table] and transcription regulation [numbers under the specific sample column]) between FUS3, ABI3 and AGL15 and the *DUF1264* genes are summarized in Table 4.1.

Briefly, FUS3 might indirectly regulate *DUF1264C* and *DUF1264D*, and might directly control *DUF1264A* and *DUF1264E*. ABI3 might directly regulate all five *DUF1264* members and AGL15 might indirectly regulate *DUF1264A* and *DUF1264E* (based on response of transcript accumulation but lack of binding to associated regulatory regions) (Table 4.1). FUS3 and ABI3 binding sites overlap for *DUF1264A* and *DUF1264E* but determining whether these two B3 domain proteins compete for the same RY motifs will require further experiments.

There are few studies about the function of DUF1264 proteins. One paper shows the *DUF1264A* and *DUF1264D* are expressed similar to *LEC2*, another embryo transcription factor gene, in that they are only “present” in the *pkl* mutant background that has been treated with uniconazole-P (a chemical inhibitor of GA biosynthesis), but “absent” in the *pkl* and wild type without uniconazole-P, and wild type with uniconazole-P samples. *PICKLE* (*PKL*) encodes a CHD3 protein, which plays a role during germination to inhibit the expression of seed-related genes. This suggests *DUF1264A*

and *DUF1264D* may exhibit PKL-dependent repression (Zhang, Rider et al. 2008). Later the same year later, another report using mass spectrometry indicated tyrosine phosphorylated DUF1264A protein is increased in response to ABA treatment (Ghelis, Bolbach et al. 2008). More recently, a member of this family in *Zea mays*, *DUF1264A*, is predominantly expressed in the scutellum during maize seed maturation and may be involved in the regulation of oil body size. The authors therefore named the protein OIL BODY ASSOCIATED PROTEIN 1 (OBAP1). Also, they reported that the OBAP1 ortholog in Arabidopsis has a reduced percentage of seeds that can complete germination compared to wild type (WT; (López-Ribera, La Paz et al. 2014). In this study, *DUF1264* genes were confirmed to be directly regulated by ABI3, FUS3 and AGL15 in different combinations that were consistent with ChIP-chip and expression microarrays results (Table 4.1). Also, it was necessary to generate higher order mutants and testing for phenotypes to establish an earlier developmental role. The (López-Ribera, La Paz et al. 2014) group indicated that *duf1264a* mutant (they named it as OBAP1) showed a much lower completion of germination percentage compared with WT. But when the same mutant line was used, *duf1264a* mutant did not show poor germination in our experiments. The *duf1264aabbCCddee* quadruple mutant has already been selected and showed a reduction in SE compared to WT control. Unfortunately, the loss-of-function allele of *DUF1264C* in the mutant collections is poor, so a CRISPR/Cas9 toolkit was used to generate a knockout in *duf1264c*. A large deletion in *DUF1264C* was generated but only in the Col WT background, and not in the *duf1264* quadruple mutant. It is possible that *duf1264* pentuple mutant is lethal in Arabidopsis, but this

needs further investigation. Interestingly, the quadruple *duf1264aaxxccddee* mutant showed increased sensitivity to ABA compared to WT (x indicates that the status of B is unclear).

## 4.3 Results and Discussion

### 4.3.1 The DUF1264 family plays a role in somatic embryogenesis

The *duf1264a, duf1264c, duf1264d,* and *duf1264e* T-DNA insertion single mutant seeds were ordered from the Arabidopsis Biological Resource Center and by request from Dr. Isobel Parkin for the *duf1264b* line from the Saskatoon SK collection (Robinson and Parkin 2009, Robinson, Tang et al. 2009). Then the higher order *duf1264* pentuple mutant was attempted by crossing the single mutant lines and genotyping the resultant progeny by PCR. To confirm the transcript level for each mutant was at least lower than the WT, the total RNA was extracted from developing seeds of each single mutant and then the transcript accumulation was determined using semi-quantitative PCR. We found the *duf1264a* and the *duf1264d* were knock-down mutations as transcripts could still be detected, although at a lower level than WT. The *duf1264b* and *duf1264e* lines appeared to be knock-out mutations as no transcripts were detected. Unfortunately, although the *duf1264c* passed the genotyping assay that proved the T-DNA insertion was present in the plants, the transcript amounts of *duf1264c* was still similar comparable to that from the WT plants based on semi-quantitative PCR result. The T-DNA insertion is located within the promoter region. Therefore, a CRISPR-Cas9 strategy was designed to knock out the *DUF1264C* gene by generating a large deletion. Before generating the pentuple mutant, SE was checked to compare the capacity of the

quadruple mutant to that of WT since the DUF1264 family were potentially directly regulated by FUS3, ABI3 and AGL15 (Table 4.1) which are sufficient to induce embryo programs. The shoot apical meristem somatic embryogenic (SAM SE) protocol was described in (Thakare, Tang et al. 2008). Experimental results of three biological replicates showed the WT had 20% of the "seedlings" in the SAM SE system with embryo production, while the quadruple *duf1264aabbCCddee* had only 12% of "seedlings" with SAM SE production. This represents a significant reduction in SE capacity (using a Student's t-test and a p-value of 0.05), suggestive of the DUF1264 family being involved in SE. Subsequently, *DUF1264A* over-expression plants were generated. Preliminary data (single biological replicate) showed an increase in SE for this over-expressing genotype (35%) compared to WT (4%), but requires replication given the unexpectedly low percentage seen (normally WT is around 20%) in the WT population.

### 4.3.2 FUSCA3, ABI3 and AGL15 regulate different members of *DUF1264* family

In the expression microarrays results of (Yamamoto, Kagaya et al. 2010), our lab found that FUS3 can positively regulate four members of the *DUF1264* family in 12 days after flowering (daf) developing seed, but *DUF1264B* is absent on the Affymetrix array used. Using qRT-PCR, the expression microarrays results could be verified and also the *DUF1264B* expression determined. An intermediate stage of seeds, between the samples in (Yamamoto, Kagaya et al.,2010) and the 15-16 daf samples used for the abi3 experiments (Chapter 2) was used for follow-up experiments (13-14 daf). The results showed that in 13-14 daf developing seeds, all *DUF1264* family members were downregulated in the *fus3* mutant compared with the WT Col (Figure 4.2). All were

significant within an independent biological replicate (3 total) with at least three technical replicates. The decreased transcript accumulation in *fus3* relative to WT Col was also significant (p-value <0.05, Student's t-test) when the biological replicates were averaged (Figure 4.2). In the expression microarrays results for ABI3 (Table 3.1), all of the *DUF1264* family members were downregulated in the 15-16 daf developing seeds in the *abi3* mutant compared with those in WT Ler (Tian, Wang et al. 2020). Here I would like to make it clear, the *fus3* mutant is in the Col background and the *abi3* is in the Ler background. Similar to FUS3 experiments, RT-qPCR confirmed the results that this *DUF1264* family was controlled by ABI3 for 13-14 daf developing seeds (Figure 4.3). This stage of seeds were used to be comparable to *fus3* and because RNA is much easier to isolate from this stage than from the 15-16 daf seeds used for the microarray.

As noted in the expression microarrays results (Table 4.1), in 10 day old shoot SAM SE tissue (Zheng et al., 2009), our lab found that the transcript accumulation of *DUF1264A*, *DUF1264C* and *DUF1264E* are insignificantly different in the *agl15agl18* mutant compared with that in WT Col. As well, over-expression of *AGL15* can up regulate the same three family members. Consequently, the *agl15agl18* double mutant was used because of the AGL18 functionally redundancy with AGL15 (Adamczyk, Lehti‐Shiu et al. 2007). The *DUF1264B* was not represented on the array and *DUF1264D* had no signal in the microarrays (Table 4.1). RT-qPCR was conducted to verify the results and quantify the missing probes on the microarray data. It was determined that all five *DUF1264* family members were up regulated in the *35S:AGL15* background compared with WT. Four biological replicates were performed, and the genes were up

regulated (*35S:AGL15*/WT) in each independent experiment but the amount of the fold change were different (Table 4.3). Even the ranking of the transcript amount between the replicates were different. Since there were additional plant hormones in the culture medium used for SAM SE, the same qRT-qPCR experiments were performed for *agl15agl18*, *35S:AGL15* and WT using 7 to 8 daf developing seeds, a stage where AGL15 transcript and protein accumulation is relatively high (Heck, Perry et al. 1995). In that stage, *DUF1264A*, *DUF1264D* and *DUF1264E* were significantly downregulated in the *agl15agl18* compared with the WT (Figure 4.4). *DUF1264*B and *DUF1264C* showed not significantly different in the *agl15agl18* compared with the WT. Surprisingly, both *DUF1264A* and *DUF1264D* were significantly repressed in the *35S:AGL15* compared to the WT (Figure 4.5). *DUF1264B, DUF1264C* and *DUF1264E* showed not significantly different in the *35S:AGL15* compared with the WT.

FUSCA3, ABI3 and AGL15 are all DNA-binding transcription factors involved in embryo development. To understand if these transcription factors directly bind the DUF1264 genes to regulate transcript accumulation, chromatin immunoprecipitation (ChIP) was performed. In the published ChIP-chip data (Table 4.1), FUSCA3 can bind to *DUF1264A* and *DUF1264E*. ABI3 can bind all family members of the *DUF1264* family, and AGL15 can bind to *DUF1264B* and *DUF1264C*. ChIP-qPCR was performed to confirm these binding results and the results were consistent (Figure 4.6 and Table 4.4), although additional biological replicates need to be done for ABI3. Finally, combining the results of the transcriptome assessment and the interaction between proteins and DNA, we confirmed (Table 4.1) and conclude that FUS3 could indirectly regulate *DUF1264B,*

*DUF1264C* and *DUF1264D* family members, and could directly control *DUF1264A* and *DUF1264E.* ABI3 might directly regulate all *DUF1264* family members. Previous reports showed both ABI3 and FUSCA3 may bind to the RY motif. All *DUF1264* family members have the RY motifs (CATGCA/TGCATG). Interestingly, there was at least one RY motif exist in the 5' UTR or 5' intergenomic region of *DUF1264B*, *DUF1264C* and *DUF1264D,* but not in *DUF1264A* and *DUF1264E.* AGL15 might indirectly control *DUF1264A*, *DUF1264D* and *DUF1264E,* and only bound to *DUF1264B* and *DUF1264C* but whether there were positive, negative or no change on transcript accumulation depended on tissue type.

### 4.3.3 Use of CRISPR-Cas9 to knock-out *DUF1264C*

Initially, I used a CRISPR-Cas9 design with a single guide RNA to attempt to make a small insertion or deletion in the target gene *DUF1264C*. Although this method is theoretically feasible and has been used successfully by others, I found it to be very inefficient and expensive since each potential mutant needs to be checked by sequencing. So, I tried a new CRISPR-Cas9 design which could knock out larger portions or even the entire *DUF1264C* gene (Yan, Chen et al. 2016).

The CRISPR-Cas9 construct was introduced by the floral dip method into the *duf1264aabbCCddee* quadruple mutant as well as into Col WT. Surprisingly, I found there were no successfully engineered plants under *duf1264aabbCCddee* quadruple mutant background. The antibiotic resistance suggested that I got successful transformation, but none of them subsequently produced the desired deletion. The construct contains a cassette for cloning the guide RNA's that will target the site of interest in the genome,

as well as a region coding for Cas9 and antibiotic resistance. An advantage of CRISPR-Cas9 is the transgenic insertion may be bred out once the guide RNA's target the gene of interest so that the resulting plants do not contain antibiotic resistance (or any other parts of the transgenic construct), but rather only the targeted and now altered gene. All the transformed plants with antibiotic resistance were determined to be false positive by PCR for the expected deletion. To eliminate any technical errors, the transformation process was repeated again but still there were no positive results for *duf1264aabbCCddee* quadruple mutant background (total of 126 plants screened). The experiment by using the same construct in the WT background was successful (two of 5 PCR screened). The DNA electrophoresis gel image is shown in Figure 4.7, and the data results are summarized in Table 4.5. The *duf1264c-1* was chosen since *duf1264c-2* might be a heterozygote (Figure 4.7, *duf1264c-2* has a blurred band in the WT position). After sequencing, the *duf1264c-1* was found to be the correct mutation and the size was 355 bp of *duf1264c-1* after CRISPR-Cas9 deletion. Sequencing for *duf1264c-2* failed and this line was not pursued.

### 4.3.4 The *duf1264aaxxccddee* quadruple mutant was hypersensitive to ABA

The *duf1264aabbCCddee* quadruple mutant and *duf1264c-1* were crossed to generate *duf1264aabbccddee* pentuple mutant. Genotyping was performed, but B (and sometimes other alleles) gave ambiguous results and so it was unclear if I had bred the *duf1264b* allele to homozygosity. To confirm the transcript level for each *DUF1264* member, the total RNA was extracted from young developing seeds and then the transcript level for each gene was checked by semi-quantitative PCR. The results showed

that the transcripts for *duf1264b, duf1264e* and *duf1264c* were knocked out; the *duf1264a* and *duf1264d* were knocked down. Then the next generation seeds of the *duf1264aabxccddee* pentuple mutant and the Col WT were used to test the ABA sensitivity. Loss-of-function of *duf1264aaxxccddee* enhances ABA sensitivity compared with the Col WT. Seeds of *duf1264aabxccddee* and WT control were allowed to complete germination on MS medium with and without ABA and seedling establishment scored by development of green cotyledons As shown in Figure 4.8, the number of seedlings with green cotyledons was reduced in *duf1264aaxxccddee* compared to WT when ABA was added to the medium (Figure 4.8A, Figure 4.8B). When 3 µmol/L of ABA was used, the *duf1264aaxxccddee* mutant developed fewer green cotyledons compared with WT at the 16th day after 3 days moist chilling to break dormancy. Twenty-four percent of the *duf1264aaxxccddee* seedlings developed green cotyledons that was significantly lower (p-value <0.05, t-test) than the 70% WT seeds with green cotyledons.

**4.4 Conclusion and Future Study**

The DUF1264 family plays an important role involved in somatic embryo development. In this study, we showed the negative impact of mutations in this family on SE, and that several key transcription factors could directly or indirectly regulate different members of the DUF1264 family. The *duf1264aabbccddee* mutant and more lines with overexpression of *DUF1264* members need to be generated and checked for the somatic production in future work.

Here, *DUF1264C* was successful removed from WT background but not from the *duf1264aabbCCddee* quadruple mutant background. The Agrobacteria transformation

process was successful, and many plants survived antibiotic selection, but all proved to be the false positive results. In another words, the CRISPR-Cas9 construct was successfully introduced into the plants but it did not remove the *DUF1264C*. The successful cases under WT background excluded the possibility that the vector itself would not work. One possible reason could be that *duf1264aabbccddee* pentuple mutant was lethal to plants. If we assume this was correct, then the next generation of *duf1264* quadruple homozygote with one family member heterozygote should has 25% dead seeds. But unfortunately, more than 90% of seeds completed germination. Even with the crosses, the genotype of B is still in question, so perhaps no C delections were obtained in the quadruple *duf1264aaccddee* mutant because a homozygous pentuple mutant would be lethal. This experiment needs to be repeated to confirm the result. Another possibility was the seeds died in the early stage of embryo developing stage, but I opened the silique when the seeds were still very young and did not find any evidence to prove this hypothesis.

Here, I showed that the *duf1264aaxxccddee* quadruple mutant was hypersensitive to ABA that suggested this family plays a crucial role involved in ABA regulation. The (López-Ribera, La Paz et al. 2014) group indicated that *duf1264a* mutant showed a much lower completion of germination percentage compared with WT. But when the same mutant line was used, *duf1264a* mutant did not show poor germination in our experiments (S. Perry, unpublished observation). It was possible that we were using fresh dry seeds while maybe they use the seeds from different time point after harvesting. As I mentioned before, the *duf1264a* and the *duf1264d* were knock down mutants. That can

be the reason why there was no phenotype for *duf1264aaxxccddee* quadruple mutant when lower concentrations of ABA (300nM and 500nM) were applied to the media. To get a clearer picture of the function for this protein family, knock out mutants of the *duf1264a* and the *duf1264d* need to be generated. Interestingly, although the transformants were all false positive for deletion under *duf1264aabbCCddee* mutant background, I found the transformation efficiency (floral dip) for *duf1264aabbCCddee* is much higher than the WT (an example is shown in Figure 4.9). More repeats and a positive control by using different constructs need to be introduced in the further research.

## 4.5 Methods

### 4.5.1 Plant Material

*Arabidopsis* Col/Ler WT, *fus3-3*, *abi3-5*, *agl15agl18*, *35S:AGL15* plants were grown as described in (Wang and Perry 2013, Tian, Wang et al. 2020). For *abi3-5* and *fus3*, siliques from heterozygous plants were opened under sterile conditions at a relatively mature stage, *abi3-5* and *fus3-3* seeds identified by green and purple color respectively and moved to germination medium to establish the homozygous mutant. The *abi3* homozygous seed remains green when heterozygotes or WT brown because ABI3 is involved in degreening of the embryo. The *fus3* homozygous mutant embryos accumulate anthocyanin that causes a purple coloration.

To establish embryonic culture tissue (ECT) for ChIP experiments, a *35Spro:AGL15* transgene was introduced by crossing with the *ABI3pro:ABI3-myc abi3-5* plants *or FUS3pro:FUS3-myc fus3-3* plants lines. Developing zygotic embryos of green bent

cotyledon stage were isolated and placed into culture as described in (Harding et al., 2003).

### 4.5.2 Somatic embryo system, ChIP-qPCR and qRT-PCR

The SAM SE production experiments were described in (Thakare, Tang et al. 2008). Briefly, mature seeds are surface sterilized and after moist chilling are placed into sterile medium containing 2,4-D in 125ml flasks. The flasks are placed on a shaker at room temperature under long day light conditions. Approximately three weeks later, the callused seedlings are scored from whether they have SE development at the shoot apical meristem (SAM) or not and % with SAM SE per flask calculated. ChIP was performed as in (Wang and Perry 2013). Experiments to confirm binding by ABI3, AGL15, and FUSCA3 (ChIP-qPCR) and response to ABI3, AGL15, and FUSCA3 accumulation (qRT-PCR) were performed as described in (Wang and Perry 2013) and Chapter 2 (Tian, Wang et al. 2020). The specific primers used for these experiments are listed in Table 4.6.

### 4.5.3 Crispr-Cas9

The vectors to generate the CRISPR-Cas9 constructs to knockout *DUF1264C* were kindly provided by Drs. Yan and Kaufmann (Yan, Chen et al. 2016). Design of the guide RNA (gRNA) was following their instruction using http://www.genome.arizona.edu/crispr/index.html. The sequences of the guide RNAs to generate a large deletion in DUF1264C were (**Seq1-1** 5'-GATTGTGGCGTGTCCGGCAAGCAG-3' **Seq2-1** 5'AAACCTGCTTGCCGGACACGCCAC 3' **Seq1-2** 5'-GATTGCCGGTTTCGCAACGAGCTC-3' **Seq2-2** 5'-AAACGAGCTCGTTGCGAAACCGGC-3'). Since the whole or at least a large portion of *DUF1264C* was desired to be removed, two

pairs of primer were designed, one was located near the 5’ end and another was located near the 3’ end of the gene. Constructs were introduced into plants using the floral dip method (Clough and Bent 1998) and transformants selected on MS medium containing 50mg/L hygromycin.

**Table 4.1** ChIP-chip (direct column) and expression microarrays results (numbers under the specific sample column) for the Duf1264 family. Direct, Yes, means that an associated regulatory region was identified by ChIP-chip experiments. Direct, no means no associated regulatory region was found. The other columns indicate the amount of transcript in the mutant genotype relative to the Col or Ler wt background that is set to 1. daf, days after flowering; NS, not significant; Absent, signal absent on array in all samples.

| AGI | Duf1264 | FUS3 data | | | ABI3 data | | AGL15 data | | |
|---|---|---|---|---|---|---|---|---|---|
| | | direct | 8daf *fus3/Col* | 12daf *fus3/Col* | direct | 15-16 daf *abi3/Ler* | direct | *agl1518/Col* | *35S:AGL15/Col* |
| At1g05510 | A | Yes | 0.31 | 0.04 | Yes | 0.06 | No | 0.8(NS) | 7.7 |
| At2g31985 | B | No | not on array | not on array | Yes | 0.03 | Yes | not on array | not on array |
| At4g18920 | C | No | 1.32 | 0.06 | Yes | 0.11 | Yes | 0.8(NS) | 3.7 |
| At5g45690 | D | No | 0.7 | 0.03 | Yes | 0.07 | No | Absent | Absent |
| At1g29680 | E | Yes | 0.11 | 0.03 | Yes | 0.15 | No | 0.8(NS) | 2.1 |

**Table 4.2** Preliminary data: an increase in somatic embryogenesis for *DUF1264A* overexpression compared to wild type control in one population.

| Genotype | WT | *DUF1264A* overexpression |
|---|---|---|
| Embryo production | 4% | 35.47% |

**Table 4.3** *DUF1264* transcript assessment under *35S:AGL15* and WT background by using RT-qPCR. Four biological replicates were performed using 10 day old SAM SE tissue.

| | | *DUF1264A* | *DUF1264B* | *DUF1264C* | *DUF1264D* | *DUF1264E* |
|---|---|---|---|---|---|---|
| Transcripts level relative to WT in *35S:AGL15* over-expression SAM SE | Replicate 1 | 1.8 | 3.86 | 3.67 | 5.08 | 3.81 |
| | Replicate 2 | 1.78 | 2.73 | 1.71 | 5.36 | 2.71 |
| | Replicate 3 | 2.676 | 237.2 | 36.827 | 1.696 | 10.685 |
| | Replicate 4 | 6.62 | 18 | 56.99 | 17.2 | 45.9 |
| | Average | 3.219 | 65.4475 | 24.79925 | 7.334 | 15.77625 |

**Table 4.4** Preliminary data. All *DUF1264* family members are bound by ABI3 in one population. *DUF1264B* and *DUF1264C* are bound by AGL15 in three replicates within themselves. Fold enrichment of regulatory regions associated with all *DUF1264* family members from ChIP using c-myc antibody and ABI3-c-myc tissue compared with no antibody control. Fold enrichment of regulatory regions associated with *DUF1264B* and *DUF1264C* from ChIP using anti-AGL15 antibody and *35S:AGL15* tissue compared with a pre-immune control (PI) control. FC: fold change compared with no-binding control.

| | *DUF1264A* | *DUF1264B* | *DUF1264C* | *DUF1264D* | *DUF1264E* |
|---|---|---|---|---|---|
| AGL15 | No | 79.84 FC | 5.09 FC | No | No |
| ABI3 | 2.65 FC | 2.32 FC | 1.68 FC | 4.28 FC | 2 FC |

**Table 4.5** Number of mutagenized (numerator) compared to number of transformants tested (denominator) for introduction of the CRISPR-Cas9 construct to knock out *DUF1264C* under *duf1264abde* quadruple mutant and WT background.

| | *duf1264abde* quadruple mutant | WT |
|---|---|---|
| Knock out efficiency | 0/126 | 2/5 |

**Table 4.6** Primers used in this study presented 5'→3'

| AGI | Gene name | Forward | Reverse |
|---|---|---|---|
| **For qRT-PCR** | | | |
| *Duf1264A* | *At1g05510* | CCGGTTGAGTCCGTTCC | ACATGATTCGAGGAGGGCTAT |
| *Duf1264B* | *At2g31985* | AGGAGGCAAAGGGATGAAAC | CTTGCTTTGAGGGCATTATTAGAG |
| *Duf1264C* | *At4g18920* | CCGTCAAGAGATGGGATCTG | GCGCAAAGCCTTTCCTAAAC |
| *Duf1264D* | *At5g45690* | AAAGATCAAGCCAGGGCTAT | TTCTTCTCCGGTCCCATAATTC |
| *Duf1264E* | *At1g29680* | CGATGGCTGACTACTGGAAG | AGGACATTGTTCACGGGAATT |
| *TUA3* | *At5g19770* | TGGTGCCCAACTGGGTTCAAAT | ACCTCTGCAACTGCTGTGTTGT |
| **For ChIP-qPCR** | | | |
| *Duf1264A* | *At1g05510* | TGTGAACACGTGTGGAGTT | GAAACCTATGACAAGCGCAAA |
| *Duf1264B* | *At2g31985* | TTTACGTACGCGCAGAGATG | CAAAGACCGGAACCTCCATAAT |
| *Duf1264C* | *At4g18920* | GAACTACTCGGTTTGCGAATTG | CCTTGCCGATCCTGATGAATA |
| *Duf1264D* | *At5g45690* | ATCATCGCCGCTCCTTTATC | CGTGACGGATCAGAGAACTTAC |
| *Duf1264E* | *At1g29680* | TGAAGGAACTCGTCGTTGAC | GCAGATGAACACTCATGTCTCTA |
| Negative Control | Non-bound region | GAACTACTCGGTTTGCGAATTG | CCTTGCCGATCCTGATGAATA |

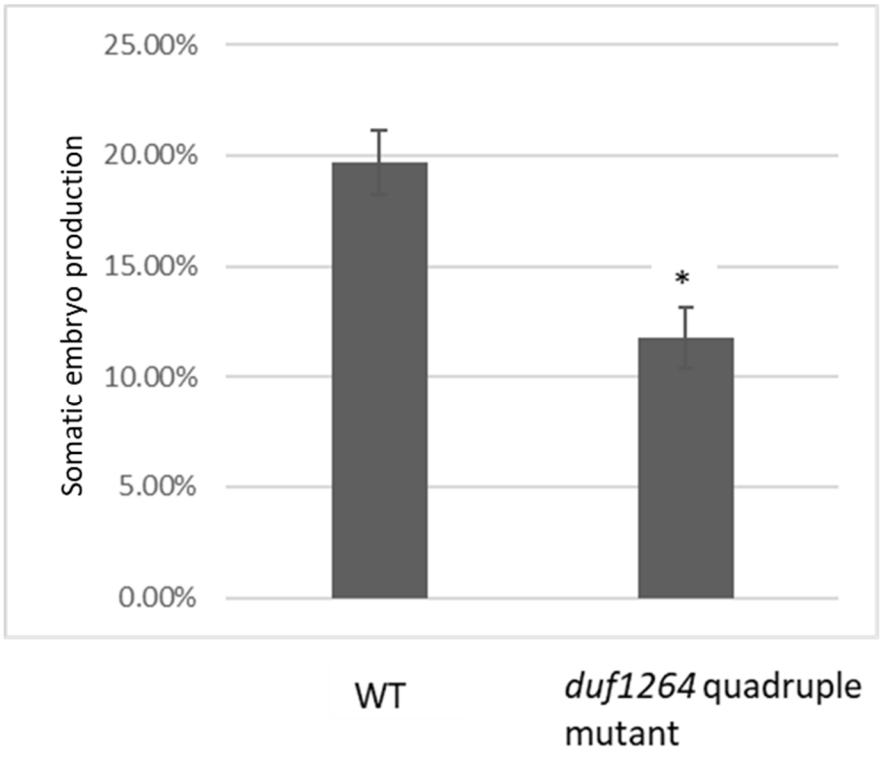
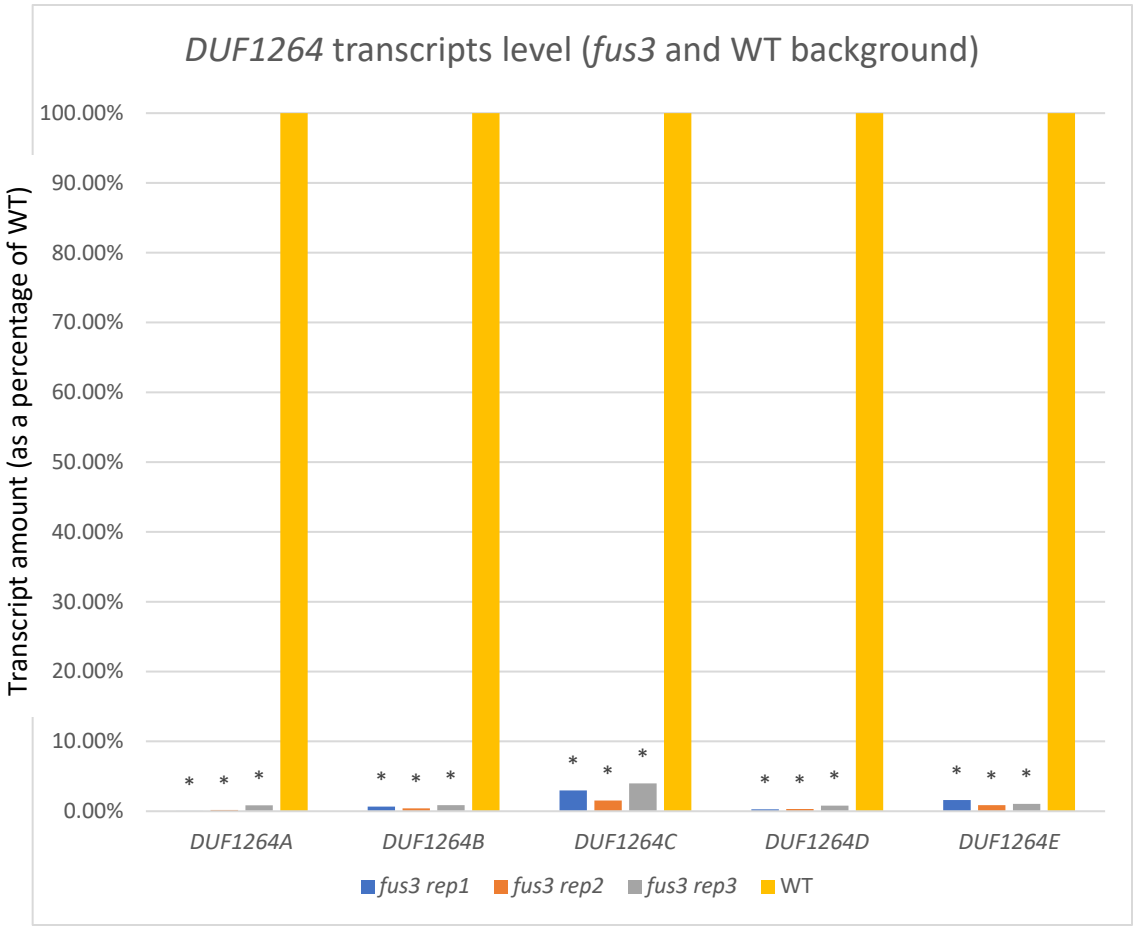
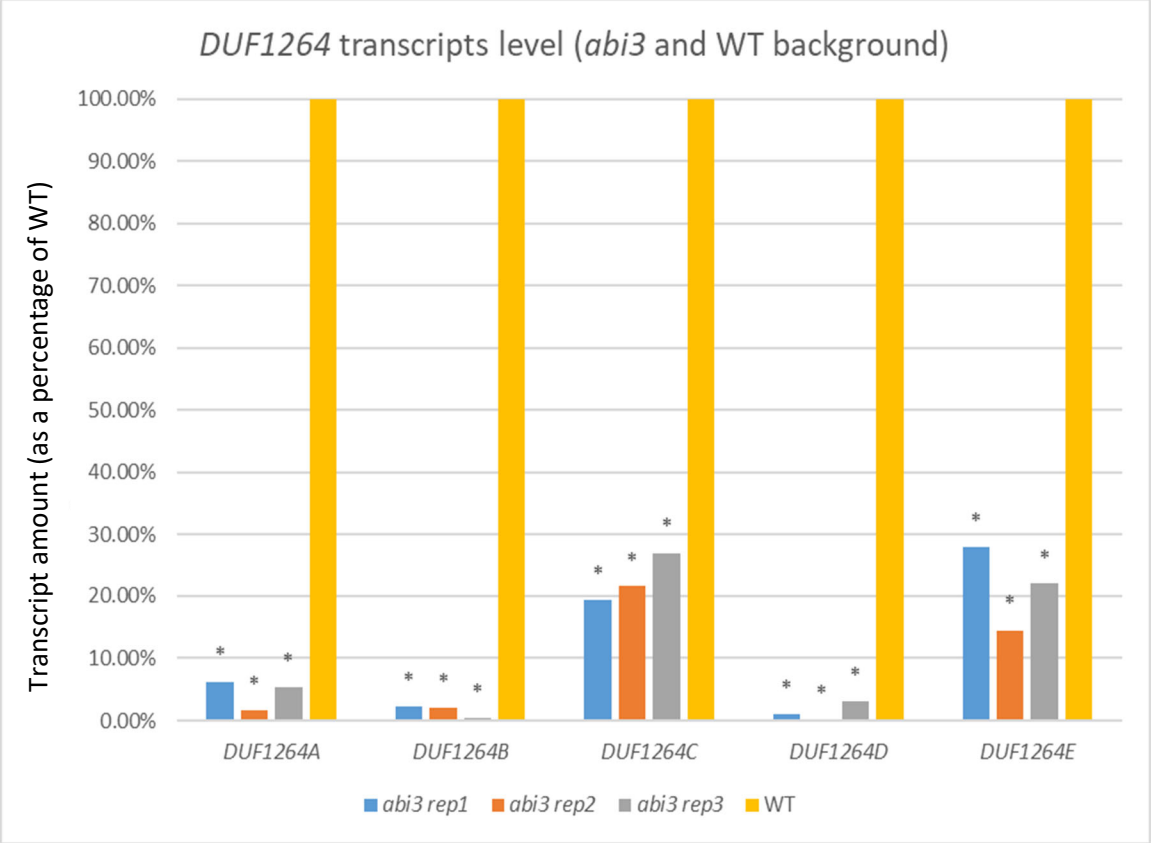



**Figure 4.1** Somatic embryo production of WT and *duf1264aabbCCddee* quadruple mutant in shoot apical meristem somatic embryogenic (SAM SE) system. The 12% of *duf1264aabbCCddee* quadruple mutant seedlings that produced embryos is significantly lower than the embryo production of 20% of WT. Data shown is from three biological replicates. * means p-value<0.05, using Student's t-test.



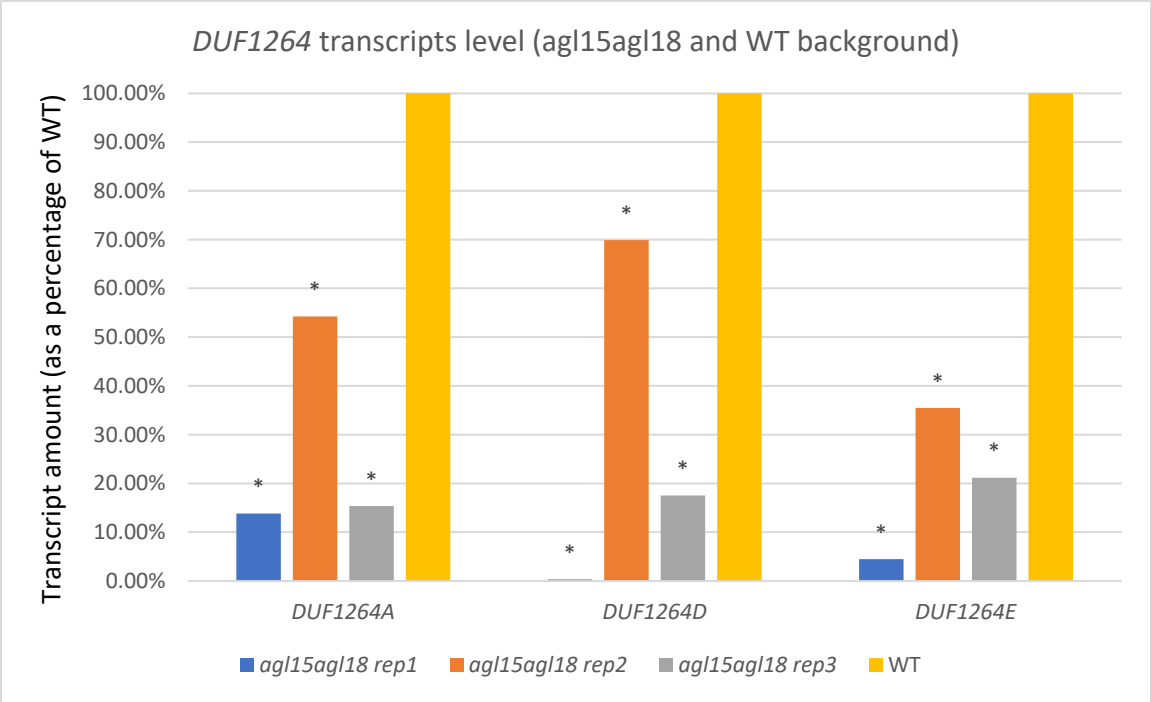

| Genotype | *DUF1264A* | *DUF1264B* | *DUF1264C* | *DUF1264D* | *DUF1264E* |
|---|---|---|---|---|---|
| *fus3* /WT (average) | 0.35% | 0.63% | 2.82% | 0.45% | 1.17% |
| TTEST, p-value (consider all replicates together) | 0.0000031 | 0.0000009 | 0.0000267 | 0.0000014 | 0.0000025 |

**Figure 4.2** *DUF1264* transcript assessment of *fus3* and WT 13-14 daf developing seeds using RT-qPCR. Each biological replicate is a different color. Data is normalized such that WT *DUF1264* transcript amount was 100%. * represents a p-value <0.05 using a Student's t-test.



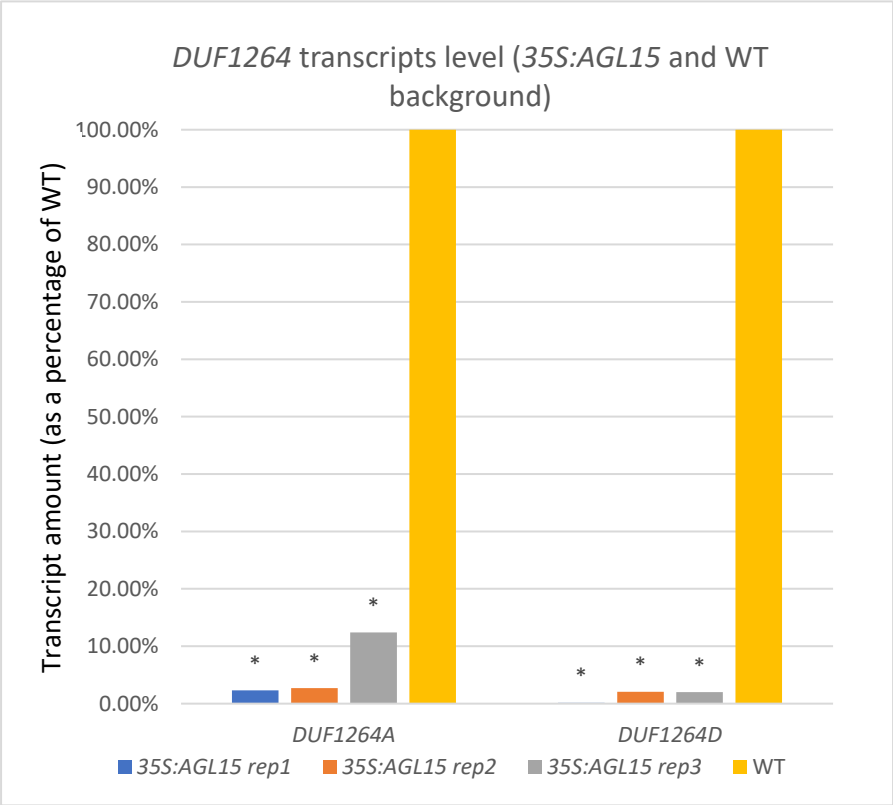

| Genotype | *DUF1264A* | *DUF1264B* | *DUF1264C* | *DUF1264D* | *DUF1264E* |
|---|---|---|---|---|---|
| *abi3* /WT (average) | 4.39% | 1.56% | 22.65% | 1.46% | 21.53% |
| TTEST, p-value (consider all replicates together) | 0.0001094 | 0.0000174 | 0.0004429 | 0.0000388 | 0.0012380 |

**Figure 4.3** *DUF1264* transcript assessment under *abi3* and WT background by using RT-qPCR (13-14 days daf developing seeds tissue). Three replications and WT were marked by different color columns. Data is normalized such that WT *DUF1264* transcripts level is 1 (100%). *, means p-value <0.05, Student's t-test.




| Genotype | *DUF1264A* | *DUF1264D* | *DUF1264E* |
|---|---|---|---|
| *agl15gl18* /WT (average) | 27.80% | 29.23% | 20.37% |
| TTEST, p-value (consider all replicates together) | 0.0159668 | 0.0387057 | 0.0062073 |

**Figure 4.4** *DUF1264* transcript assessment under *agl15ag18* and WT background by using RT-qPCR (7 to 8 daf developing seed). Three replications and WT were marked by different color columns. Data is normalized such that WT *DUF1264* transcripts level is 1 (100%). *, means p-value <0.05, Student's t-test.




| Genotype | *DUF1264A* | *DUF1264D* |
|---|---|---|
| *35S:AGL15* /WT (average) | 5.80% | 1.39% |
| TTEST, p-value (consider all replicates together) | 0.0006118 | 0.0000205 |

**Figure 4.5** *DUF1264A* and *DUF1264D* transcriptomes assessment under *35S:AGL15* and WT background by using RT-qPCR (7 to 8 daf developing seed). Three replications and WT were marked by different color columns. Assumed the WT *DUF1264* transcripts level is 1 (100%). *, means p-value <0.05, Student's t-test.

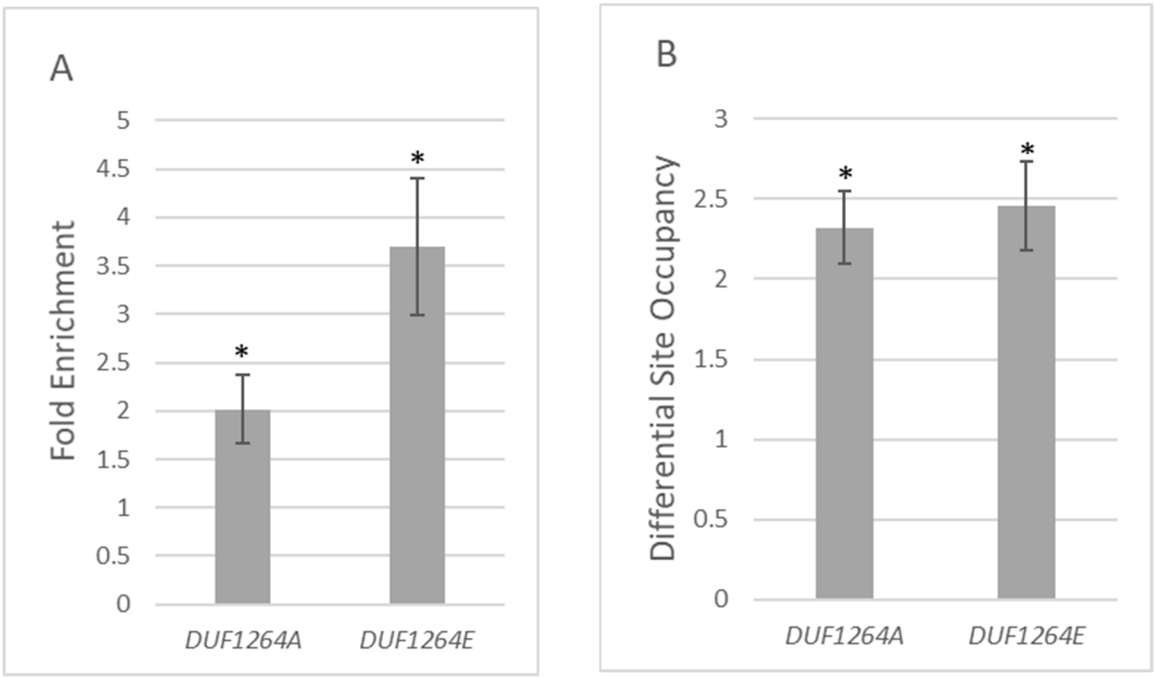



**Figure 4.6** The regulatory regions of *DUF1264A* and *DUF1264E* are bound by FUSCA3. A, Fold enrichment of regulatory regions associated with *DUF1264A* and *DUF1264E* from ChIP using c-myc antibody and FUS3-c-myc tissue compared with the control without the antibody. B, Differential site occupancy measures the amount of the suspected target recovered compared with a non-bound control fragment in the same immunoprecipitation. Significant differences for the suspected targets with antibody compared with no antibody are shown with asterisks: *$P < 0.05$.

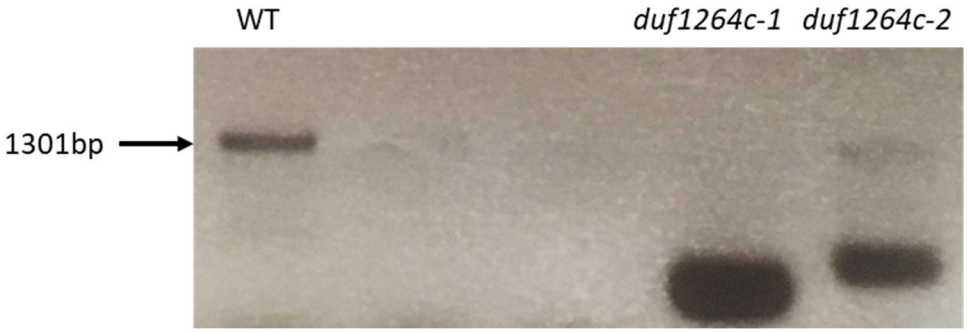



**Figure 4.7** DNA electrophoresis gel image for two successful *DUF1264C* engineered plants. The original size of WT is 1301bp, and both engineered plants are below 500 bp.

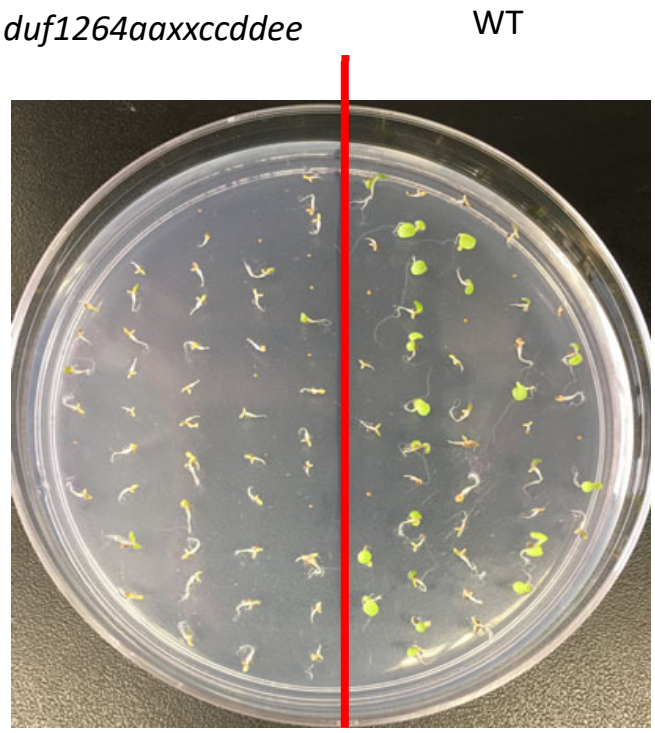



**Figure 4.8A** An example of *duf1264aaxxccddee* and WT seedling growth in response to 3 µM ABA. Left half of the plate is *duf1264aaxxccddee* mutant and right half of the plate is WT. Picture was taken at 16 days after three days moist chilling.

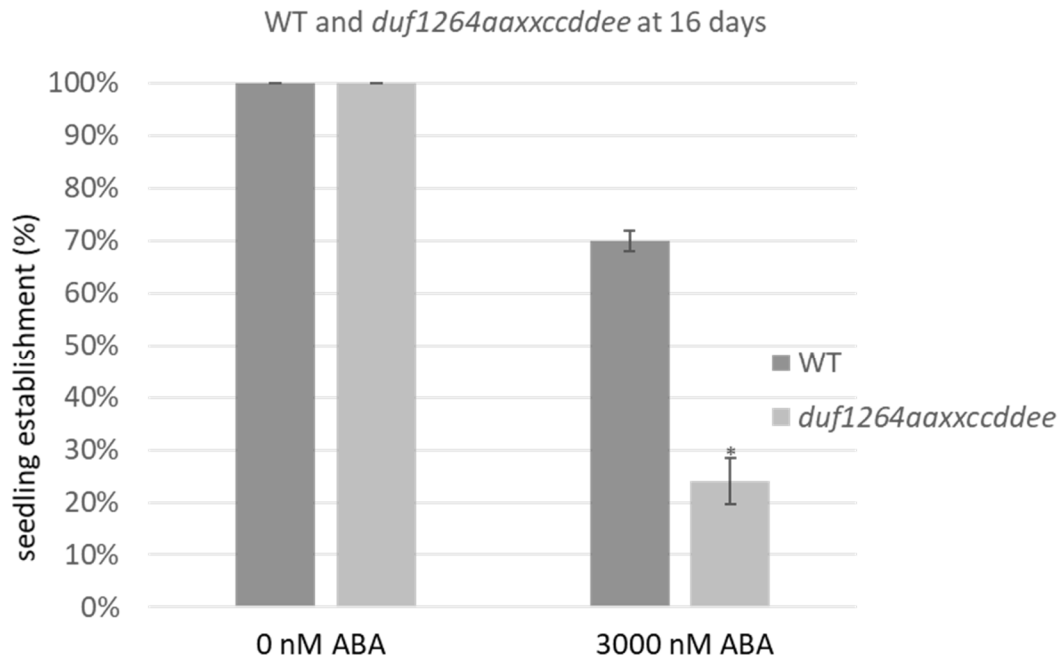



**Figure 4.8B** Seedling establishment of the *duf1264aaxxccddee* and WT in response to ABA. Seeds of *duf1264aabxccddee* and WT were directly grown on MS medium supplemented with 0 and 3 µmol concentration of ABA, and seeds moist chilled for 3 days. Green cotyledon percentages were recorded at 16 days. Each value is the mean of at least 50 seeds from 3 independent experiments. Student's t-test was used to determine the significance between Col WT and *duf1264aaxxccddee,* * means p-value

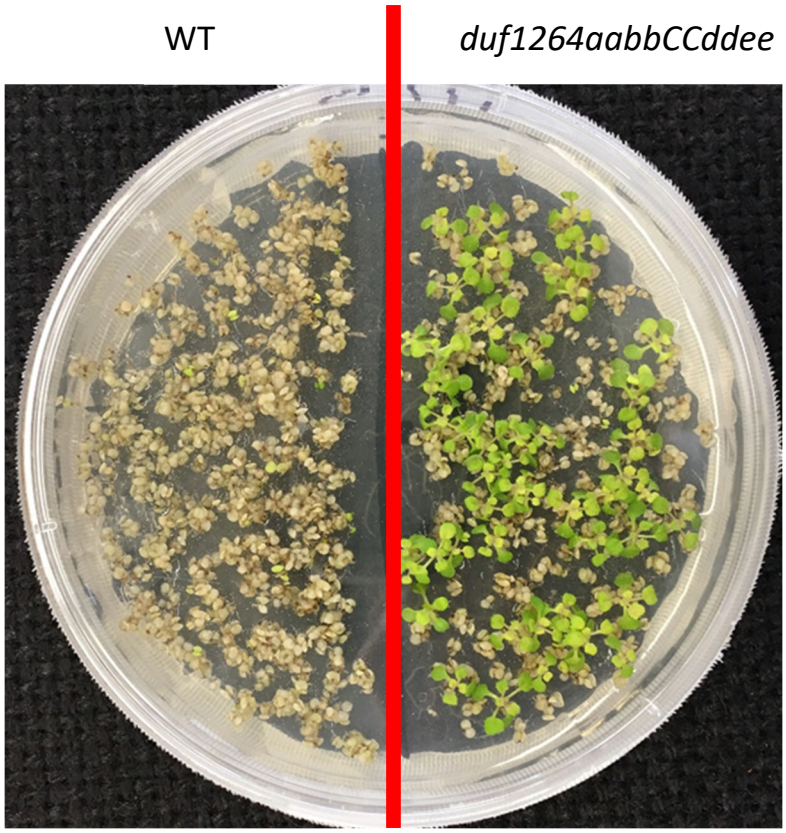



**Figure 4.9** An example of *duf1264aabbCCddee* and WT transformation efficiency (floral dip). The Left half of plate is WT and the right half is *duf1264aabbCCddee* quadruple mutant transformed with Crispr-Cas9 construct for knocking out the *DUF1264C*. hygromycin 50mg/L was used for antibiotic selection.

# Chapter 5 Comparison of datasets and conclusion

## 5.1 Abstract

*AGAMOUS-Like15* (*AGL15*) and its ortholog from *Glycine max*, *GmAGL15*, are MADS domain transcription factor genes that are preferentially, but not exclusively, expressed during seed development. Ectopic expression of *AGL15* dramatically promotes secondary embryo production from zygotic embryo explants without the addition of exogenous hormones. Also, *35S:GmAGL15* was able to promote somatic embryogenesis (SE) from cotyledon explants from soybean placed into culture on medium with 2,4-D. By identifying the direct and indirect downstream regulated genes, it may be possible to understand how (Gm)AGL15 promotes SE and provide methods to improve crops regeneration through SE. Here, I performed RNA-seq and combined results with previous (our lab) RNA-microarray to assess transcript accumulation from genes that showed significant response to over-expression of *GmAGL15* with or without 2,4-D at different time points. Also, I performed RNA-seq and ChIP-seq and combined previous (our lab) RNA-microarray and ChIP-microarray for global identification of targets of *Arabidopsis* AGL15. *Arabidopsis thaliana* SIN3A ASSOCIATED POLYPEPTIDE P18 (AtSAP18) is a co-regulator that can interact with AGL15. Here, I found 17 genes that were directly bound by and responsive AGL15 and at same time were direct associated with and responsive to AtSAP18.

## 5.2 Introduction

Somatic embryogenesis (SE) is an artificial process by which single or multiple somatic cells de-, then re-differentiate to form an embryo and subsequently a plant

(Feher 2019). Early zygotic embryogenesis is difficult to study due to the inaccessibility of higher plant embryos during this stage in that the limited numbers of cells are embedded in layers of maternal tissues. SE provides an alternative system, which has been used as a model to study zygotic embryogenesis. Also, SE is one of the important methods to regenerate plants after transformation even though many crops do not regenerate well. So, it is important to understand the mechanism of SE, especially the early stage of SE that is not well understood. SE may be formed by a direct or indirect pathway and depends on many aspects such as age of the explant, cultivar, and induction medium. Normally, a SE system requires the synthetic auxin, 2,4-dichlorophenoxyacetic acid (2,4-D) to induce the SE, but also SE has occurred without the addition of growth regulator in some cases ;(Horstman, Li et al. 2017).

Many genes, including *LEC1, LEC2, FUSCA3,* and *WUSCHEL*, have been identified to be sufficient to promote SE; in that ectopic expression of such genes can induce SE (Lotan, Ohto et al. 1998, Stone, Kwong et al. 2001, Gazzarrini, Tsuchiya et al. 2004). The details are described in Chapter 1, Literature Review. Over-expression of AGL15 can dramatically promote secondary embryo from zygotic embryo explants, whereas loss-of-function of AGL15 can decrease the frequency of SE (Harding, Tang et al. 2003, Thakare, Tang et al. 2008). To understand the role of AGL15 in crops, our lab previously isolated an ortholog of AGL15 from *Glycine max* (soybean) (GmAGL15). And found ectopic expression can enhance SE in soybean (Thakare, Tang et al. 2008). Also, to extend the understanding of the mechanism by which GmAGL15 promotes SE, our lab previously assessed the transcriptome using the Affymetrix Soybean Genome Array (Zheng and Perry 2014).

Since the microarrays previously used do not cover the entire soybean genome, here I performed the RNA-seq for samples from the same *AGL15*-overexpression line (8981) and wild type (WT)(Jack) that was previously used, including cotyledon explants from embryos of WT(Jack) and *35Spro:GmAGL15* at 0, 3, and 7 days on the medium supplied with 2,4-D (D40 medium). I combined two datasets and found that 149, 147, and 37 genes were induced by over-expression *GmAGL15* compared with WT Jack at 0, 3, and 7 days after placement in culture, (dac), respectively. For the same comparison, 171, 198 and 50 genes were found to be reduced in the *35S:GmAGL15* compared with WT Jack at the different time points. I also combined previous ChIP-microarrays and RNA-microarrays with ChIP-seq and RNA-seq data in this study for AGL15 in *Arabidopsis*, which resulted in a list of potential genes that might be involved in embryo development to support future experiments. Finally, I found the overlap with genes regulated by both AGL15 and SAP18, which might be involved in the transition phase before and after seeds complete germination. This chapter summarizes results from comparing these large datasets.

## 5.3 Results and Discussion

### 5.3.1 Comparison of the *GmAGL15* accumulation

The cotyledon explants from embryos of WT Jack and *35Spro:GmAGL15* were collected at 0, 3, and 7 days on D40 (SE induction) medium (days after placement in culture, dac), frozen in nitrogen, and the RNA extracted and cleaned with the QIAGEN RNeasy Plant Mini Kit. The RNA samples from such stages of the same line *35Spro:GmAGL15* (8981) and WT Jack were used to assess transcriptomes using the

Affymetrix Soybean Genome Array and RT-qPCR in the previous report (Zheng and Perry 2014) and RNA-sequencing in this study. Unfortunately, the *GmAGL15* itself is not in the previous Affymetrix Array data (it is not represented on the array), so the RT-qPCR was performed to assess the transcript abundance of *GmAGL15* and confirm over accumulation. The results showed the *GmAGL15* amounts in *35Spro:GmAGL15* relative to WT Jack in both previous populations prepared for RT-qPCR and the current RNA-sequencing were very close among all three stages (Figure 5.1). The results confirmed that the *GmAGL15* was successfully over-expressed on the day that (0 dac), and after (3 and 7 dac), the cotyledon explants from embryos are placed on D40. With time in culture, there was a larger increase in the amount of *GmAGL15* transcript in the overexpression sample compared to WT Jack.

### 5.3.2 GmAGL15-Responsive Genes

Our lab previously reported genes that may be responsive to GmAGL15 accumulation based on the microarray experiments. In that report our lab found that compared with the WT, the number of genes with significantly different transcript levels in *35Spro:GmAGL15* tissue is largest at 0 dac and dramatically smaller at 3 and 7 dac on D40. But in this study, when the RNA-sequencing was performed on the samples from same *35Spro:GmAGL15* (8981) soybean line, surprisingly I found that the largest number of genes was at 3 dac, and the smallest is at 0 dac on D40 (Table 5.1). The difference between two results may be caused by several reasons. Firstly, although the samples used were from the same soybean lines and the transcript levels of *GmAGL15* were similar in the corresponding samples (Figure 5.1), differences still existed. Moreover, the different

technical methods were used, in which the microarray data covers 37593 genes (Tian, Wang et al. 2020) and the RNA-seq included 56044 genes in this study. Obviously, the RNA-seq data might has the potential to discover more genes since it contained much more genes compared with the previously used microarray data. Also, in this study I prepared three biological replicates to generate the high throughput sequencing while the microarray data are duplicates, which made the RNA-seq dataset relatively ?more reliable. In short, the results from RNA-seq suggested that the greatest number of genes were activated by the accumulation of *GmAGL15* and the addition of 2,4-D together at relatively early stage of SE (3 dac). The number of genes that were activated by over-expression of *GmAGL15* itself (0 dac) or the later stage of SE on D40 (7 dac) were lower (Table 5.1).

In the previous study, our lab found that there is a large overlap (35%) between the genes that are significantly ($p < 0.05$) up-regulated by GmAGL15 (1.5 fold change or greater) and the genes that are also significantly expressed in WT Jack at 3 dac compared with 0 dac using the same cutoffs. Also, of the significant down-regulated genes by the accumulation of GmAGL15 at 0 dac, 43% are significantly repressed in WT Jack at 3 dac compared with 0 dac (Figure 5.2A). To intuitively observe the difference between the results of previous microarray and RNA-seq in this study for the same comparisons mentioned above, a similar figure was drawn (Figure 5.2B). The figure was generated based on the RNA-seq results using same cutoffs ($p <0.05$, absolute fold change > 1.5). Because the increasing coverage of RNA-seq method I used in this study compared with the microarray used before, I expected to discover more potential genes in each category.

Here I would like to make it clear, it is not the RNA-seq, the method itself, that can cover more genes compared with microarrays. The microarrays can also cover whole genomes but what our lab used before cannot cover whole genome due the limited number of specific probes. For RNA-seq, the sequencing coverage depth could easily be increased to identify lower abundance transcript, also the RNA-seq used in this study could cover the whole soybean genome. Surprisingly, except the number of genes that showed significantly increased transcript accumulation at 3 days compared to 0 day WT Jack that increased from 3731 (Microarrays) to 4726 (RNA-seq), the number of genes had decreased in other comparisons between these two platforms (Figure 5.2). This might be caused by the numbers of experiments for both methods was different, specifically the microarray data were duplicates while the RNA-seq data were triplicates at each time point per genotype. To better understand the difference between the experimental data of these two different methods, I compared the lists of gene with significant differences of *35Spro:GmAGL15* compared with WT Jack at each time point in their respective experiments (Table 5.2). In the previous microarrays, many potential targets had no soybean match identified and also some genes with Glyma1.1 identified ID have no current Glyma2.0 corresponding identified ID, so I removed those targets from the comparisons. In short, by combining the two high-throughput sequencing results, I found 149 genes were significantly expressed and 171 genes were significantly repressed by over-expression of *GmAGL15* compared with WT Jack at 0 dac. Similar numbers of genes were found to be significantly different in *35Spro:GmAGL15* tissue compared with WT Jack at 3 dac. Much lower number of genes were found to be significantly different in

*35Spro:GmAGL15* tissue compared with WT Jack at a later stage on D40. Since these potential targets were identified by both methods, these will be more reliable to pursue in future work. Also, to narrow the lists to find the potential genes that can promote SE, I compared the data to find the genes that were consistently increased or decreased by over-expression *GmAGL15* from 0 dac to 7 dac. Unfortunately, no gene was in such list when the both RNA-seq and microarray results are combined. There were some overlaps if I only focused on the genes those were significantly up-regulated or down-regulated by ectopic expression of *GmAGL15* at 0 dac and 3 dac or 3 dac and 7 dac. For example, the Glyma.09G274000 were significantly up regulated by *35Spro:GmAGL15* compared WT at both 0 dac and 3 dac time points. The ortholog of Glyma.09G274000 in Arabidopsis and *Medicago sativa* L. is WRKY DNA-binding protein 70 (WRKY70), which is involved in jasmonic acid (JA)-mediated signal pathway and plays a role as a repressor of JA-responsive genes (Li, Brader et al. 2004). There is limited literature about JA impacts on SE. One bioinformatic study suggests that potential effects of JA on SE in *Arabidopsis* (Elhiti, Stasolla et al. 2013) and others studies indicate applying the exogenous MeJA may inhibit SE of *Medicago sativa* L (Ruduś, Kępczyńska et al. 2006). Also, JA is produced in *Medicago sativa* L. callus during the whole process of indirect SE (Ruduś, Weiler et al. 2009). Glyma.18G203500 was also significantly expressed by *35Spro:GmAGL15* compared WT Jack at both 0 and 3 days on D40, but not the 7 dac time point. The *Arabidopsis* ortholog of Glyma.18G203500 is LATE EMBRYOGENESIS ABUNDANT 1 (Em1) , which is normally expressed only in embryos and only during the final maturation phase of embryo

development (Wise 2003, Manfre, Lanni et al. 2006). The complete lists are shown in the Table 5.3.

### 5.3.3 Global Identification of Targets of AGL15 in Arabidopsis

Our lab previously identified the targets of AGL15 by using a ChIP-chip approach and Affymetrix tiling arrays for *Arabidopsis thaliana* (Zheng, Ren et al. 2009). Here the RNA-seq and ChIP-seq were used to verify the previously found targets. By combining the ChIP-chip or ChIP-seq and expression array or RNA-seq results can distinguish genes that are directly bound by AGL15 to genes that are only regulated but not bound by AGL15. Our lab previously used 10 day SAM SE tissue for Affymetrix tiling arrays and culture tissue for ChIP (Zheng, Ren et al. 2009). This time the ChIP experiment still remained the same, in which I used the culture tissue, but the 7 to 8 day young developing seeds were used in RNA-seq instead of SAM SE tissue. Here I combined all four results to find the overall direct and indirect targets for AGL15. Only the genes that were identified by all four dataset results were considered direct responsive targets of AGL15. Genes with transcript accumulation significantly changed in *agl15agl18* or *35S:AGL15* in both SAM SE and seed tissue but not bound by AGL15 were indirect targets of AGL15. As shown in Figure 5.3, total 74 genes were expressed AGL15 in both SAM SE and seed tissue. Ten of 74 (11.9%) were also to be the direct targets of AGL15. On another hand, a total of 63 genes were repressed by AGL15, in which 19 (23.2%) were also bound by AGL15. All the results mentioned above used the threshold p-value less than 0.05 and for the assessment of transcripts, absolute fold change above 2 was applied to define the genes that expressed or repressed by AGL15. Since I have both mutant *agl15agl18* and over-expression of

*AGL15* compared with WT Col, genes regulated by AGL15 is at least one comparison (*agl15agl18*/Col; or *35S:AGL15*) and no significant change for the other comparison, were included in the responsive list. In short, combining the four high-throughput sequencing results together provided a short, but reliable gene list (total 29) that represent direct regulated targets by AGL15 that could be pursued in future experiments.

### 5.3.4 Overlap with Genes Regulated by AGL15 and SAP18

A previous report suggested that AGL15 might interact with *Arabidopsis thaliana* SIN3A ASSOCIATED POLYPEPTIDE P18 (AtSAP18) in yeast 2-hybrid and gel shift assay experiments (Hill, Wang et al. 2008). Interestingly, unlike *AGL15* that is mostly only expressed in the seeds, *SAP18* transcripts are detected not only in seeds but also leaves including cotyledons (Nakabayashi, Okamoto et al. 2005, Schmid, Davison et al. 2005, Klepikova, Kasianov et al. 2016). Currently, AtSAP18 is known to involved in stress-response in Arabidopsis (Song and Galbraith 2006)(Chapter 3 of this thesis), but there is no evidence that it may be involved in embryo development or the transition phase before and after seeds complete germination. So, I compared the ChIP-seq in this study for AGL15 to the SAP18 data reported in the Chapter 3 of this dissertation. Both data were generated from ECT tissue so differences in binding were not due to seed age. For AGL15, 9858 binding genes were reported, whereas for SAP18, this numbers 3047 using the same cutoffs in CLC Genomics benchwork. Since the SAP18 is a co-regulator and not a transcription factor, here I defined the 3047 genes were associated with SAP18. A little more than eleven hundred (1177) genes were bound/associated by both AGL15 and SAP18. GO term analysis showed these 1177 genes were overrepresented for the

categories “response to hormone” (FDR 2.67E-02), “response to stress” (4.84E-02), and “response to ethylene”(FDR 6.06E-03). Seventeen genes were not only bound by AGL15 and associated with SAP18 but were also induced/reduced in response to them by using $p < 0.05$ and 2.0 fold change cutoffs. Those genes might be the potential candidates that are involved in the transition from embryo to seedling. The complete list of these 17 genes is shown in the Table 5.4.

**5.4 Conclusion**

Seeds make up approximately 70% of the human diet directly so understanding regulatory mechanisms to generate a seed and the embryo it contains is fundamentally important. I performed necessary follow-up work to verify global direct and indirect ABI3 target genes, determine roles in embryogenesis and analyzed the data for the paper. ABI3 could directly induce and repress its target genes’ transcript accumulation and some intriguing differences exist in *cis* motifs between these groups of genes. I found that ABI3 also regulated some genes encoding important miRNAs and, in some cases, switched between inducing to repressing expression depending on developmental stage. I also found ABI3 directly regulated all five *DUF1264* members in Arabidopsis, while two other seed transcription factors FUS3 and AGL15 directly controlled subsets of genes in this family. Here, the direct association of the TFs with these genes and effect on transcript accumulation was verified. Also, higher order mutants were generated. Duf1264 quadruple mutant showed a reduction in SE compared to WT control. Duf1264 putative pentuple mutant showed a hypersensitive response in seedlings to ABA compared to WT. In an associated project, an Arabidopsis protein called SIN3A ASSOCIATED POLYPEPTIDE

18 (AtSAP18) was investigated. SAP18 is a transcriptional co-regulator that is a component of histone deacetylase (HDAC) complexes which interacts with a TF of interest in the lab, AGL15 to control embryogenesis. A list of genes was found to be direct bound by AGL15 and at same time were direct associated with SAP18. Also, a new phenotype of a loss-of-function mutant *sap18* in that the mutant was hypersensitive to ABA treatment compared to Col WT, suggesting an important role of SAP18 in modulation of ABA response. Finally, the global targets of AGL15 were identified by combining previous RNA microarrays and ChIP microarrays method with RNA-seq and ChIP-seq in this study. Also, some of these regulatory networks were investigated in the important crop plant, *Glycine max*, and a list of potential genes that may be involved in SE was discovered.

**Table 5.1**. The number of genes with significantly different transcript level in *35Spro:GmAGL15* tissue compared with wild type Jack. Blue cells: RNA-seq data, white cells: Microarray data. Cutoffs used were p-value <0.05, fold change greater than 1.5 or less than -1.5 for both data. The microarray data shown are duplicates while the RNA-seq data shown are triplicates at each time point per genotype. Absolute fold change: absFC.

| Comparison | Genes numbers (p-value < or =0.05, absFC>1.5) | Up regulated | Down regulated |
|---|---|---|---|
| RNA-seq, ***35S:GmAGL15*** 0 day vs Jack 0 day | 2225 | 825 | 1400 |
| Microarray, ***35S:GmAGL15*** 0 day vs Jack 0 day | 4154 | 1916 | 2238 |
| RNA-seq, ***35S:GmAGL15*** 3 days vs Jack 3 days | 3290 | 1315 | 1975 |
| Microarray, ***35S:GmAGL15*** 3 days vs Jack 3 days | 2308 | 1244 | 1064 |
| RNA-seq, ***35S:GmAGL15*** 7 days vs Jack 7 days | 2581 | 1247 | 1284 |
| Microarray***, 35S:GmAGL15*** 7 days vs Jack 7 days | 664 | 174 | 490 |

**Table 5.2** The number of genes with significant differences in transcript accumulation in *35Spro:GmAGL15* tissue compared with wild type Jack from RNA-seq and microarray data. In addition, the number of genes with differential transcript accumulation shared in both datasets is shown in red. The gene numbers shown from the microarray data do not include the features on the array that did not have a Glyma 2.0 soybean associated ID.

| **Comparisons** | **Gene numbers (up regulated)** | **Gene numbers (down regulated)** |
|---|---|---|
| RNA-seq, *35S:GmAGL15* 0 day vs Jack 0 day | 825 | 1400 |
| Microarray, 35S:GmAGL15 0 day vs Jack 0 day | 1137 | 1425 |
| Overlap between RNA-seq and Microarray 35S:GmAGL15 0 day vs Jack 0 day | 149 | 171 |
| RNA-seq, 35S:GmAGL15 3 days vs Jack 3 days | 1315 | 1975 |
| Microarray, 35S:GmAGL15 3 days vs Jack 3 days | 791 | 673 |
| Overlap between RNA-seq and Microarray 35S:GmAGL15 3 days vs Jack 3 days | 147 | 198 |
| RNA-seq, 35S:GmAGL15 7 days vs Jack 7 days | 1247 | 1284 |
| Microarray, 35S:GmAGL15 7 days vs Jack 7 days | 106 | 335 |
| Overlap between RNA-seq and Microarray 35S:GmAGL15 7 days vs Jack 7 days | 37 | 50 |

**Table 5.3** Genes that were consistently and significantly up-regulated or down-regulated by ectopic expression of *GmAGL15* at two out of three time points based on the agreement of prior microarray data and RNA-seq data in this study. Threshold: p-value <0.05, absolute fold change > 1.5; Not significant: NS.

| **Gmax 2.0 Primary Protein ID** | **Significant up or down-regulated of *35S:GmAGL15* vs WT Jack at different days after placement in culture (dac)** | | |
|---|---|---|---|
| | 0 | 3 | 7 |
| Glyma.09G274000 | up | up | NS |
| Glyma.09G246600 | up | up | NS |
| Glyma.03G205500 | up | up | NS |
| Glyma.09G285900 | up | up | NS |
| Glyma.16G169700 | NS | up | up |
| Glyma.09G210900 | NS | up | up |
| Glyma.08G105400 | NS | up | up |
| Glyma.09G285700 | NS | up | up |
| Glyma.09G258600 | NS | up | up |
| Glyma.18G203500 | NS | up | up |
| Glyma.09G248600 | NS | up | up |
| Glyma.09G254500 | NS | up | up |
| Glyma.09G259200 | NS | up | up |
| Glyma.09G275900 | NS | up | up |
| Glyma.09G253100 | NS | up | up |
| Glyma.06G232300 | NS | up | up |
| Glyma.09G223700 | NS | up | up |
| Glyma.18G152900 | down | down | NS |
| Glyma.04G222800 | down | down | NS |
| Glyma.13G137200 | down | down | NS |
| Glyma.13G266600 | NS | down | down |
| Glyma.04G015000 | NS | down | down |
| Glyma.19G238000 | NS | down | down |
| Glyma.08G013200 | NS | down | down |
| Glyma.13G059500 | NS | down | down |

**Table 5.3** (continued)

| | | | |
|---|---|---|---|
| Glyma.05G124900 | NS | down | down |
| Glyma.10G125000 | NS | down | down |
| Glyma.17G261500 | NS | down | down |
| Glyma.09G022500 | NS | down | down |
| Glyma.13G175300 | NS | down | down |
| Glyma.17G234700 | NS | down | down |
| Glyma.02G045100 | down | NS | down |

**Table 5.4.** Genes were found to be directly bound by AGL15 and as well as directly associated with SAP18. Threshold: p-value<0.05, absolute fold change > 2; Not significant: NS.

| AGI | TAIR description | *sap18*/Col WT | *agl15agl18*/Col WT | *35S:AGL15*/Col WT |
|---|---|---|---|---|
| AT1G22590 | AGAMOUS-LIKE 87, AGL87 | -2.04 | NS | -2.10 |
| AT1G73190 | ALPHA-TIP, ALPHA-TONOPLAST INTRINSIC PROTEIN, TIP3;1 | 3.28 | NS | -19.53 |
| AT1G78850 | APPLE DOMAIN LECTIN-1, GAL1, MANNOSE BINDING LECTIN1, MBL1 | 2.27 | NS | -4.19 |
| AT2G02000 | GAD3, GLUTAMATE DECARBOXYLASE 3 | 3.66 | NS | -9.12 |
| AT2G24762 | ATGDU4, GDU4, GLUTAMINE DUMPER 4 | 2.04 | NS | 2.03 |
| AT2G27880 | AGO5, ARGONAUTE 5, ATAGO5 | -2.05 | NS | -5.23 |
| AT2G32510 | MAPKKK17, MITOGEN-ACTIVATED PROTEIN KINASE KINASE KINASE 17 | 2.08 | NS | -12.64 |
| AT3G22120 | CELL WALL-PLASMA MEMBRANE LINKER PROTEIN HOMOLOG (CWLP) | 2.57 | NS | -4.33 |
| AT3G51238 | Natural antisense transcript overlaps with AT3G51240 | -4.45 | NS | -8.98 |
| AT3G53040 | LATE EMBRYOGENESIS ABUNDANT PROTEIN, PUTATIVE / LEA protein | 2.46 | NS | -37.01 |
| AT1G15960 | ATNRAMP6, NRAMP METAL ION TRANSPORTER 6, NRAMP6 | -2.49 | 2.49 | NS |

**Table 5.4** (continued)

| | | | | |
|---|---|---|---|---|
| AT1G36180 | ACC2, ACETYL-COA CARBOXYLASE 2 | 2.00 | 2.99 | NS |
| AT1G53480 | MRD1, MTO 1 RESPONDING DOWN 1 | 8.18 | 43.20 | NS |
| AT2G36640 | ATECP63, ECP63, EMBRYONIC CELL PROTEIN 63 | 2.44 | 2.57 | NS |
| AT3G22370 | ALTERNATIVE OXIDASE 1A, AOX1A, ATAOX1A, ATHSR3, HSR3, HYPER-SENSITIVITY-RELATED 3 | 2.67 | 3.60 | NS |
| AT5G07330 | NFU1 iron-sulfur cluster protein | 2.33 | -18.08 | NS |
| AT5G55270 | ATDOB14, DUF295 ORGANELLAR B 14 | 12.53 | 19.41 | NS |

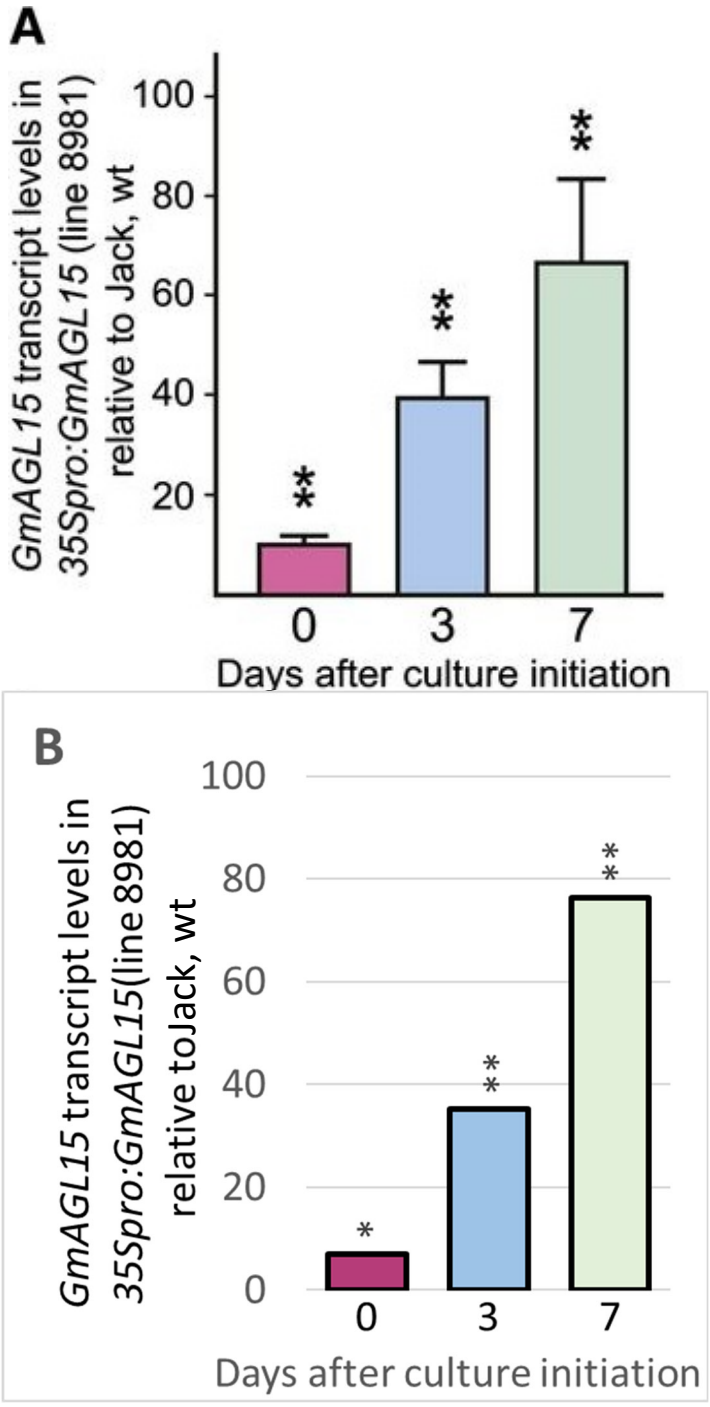

A
GmAGL15 transcript levels in 35Spro:GmAGL15 (line 8981) relative to Jack, wt
100
80
60
40
20
0
3
7
Days after culture initiation
B
GmAGL15 transcript levels in 35Spro:GmAGL15(line 8981) relative toJack, wt
100
80
60
40
20
0
3
7
Days after culture initiation

**Figure 5.1**. Transcript accumulation from *GmAGL15* in isolated immature cotyledon explants at 0, 3 and 7 dac on D40 medium. A, RT-qPCR result from a prior paper is shown (Zheng and Perry 2014). B, RNA-sequencing results in this study. Data shown are means for three biological replicates. Asterisks mean significant difference at **P < 0.01 and *P < 0.05.

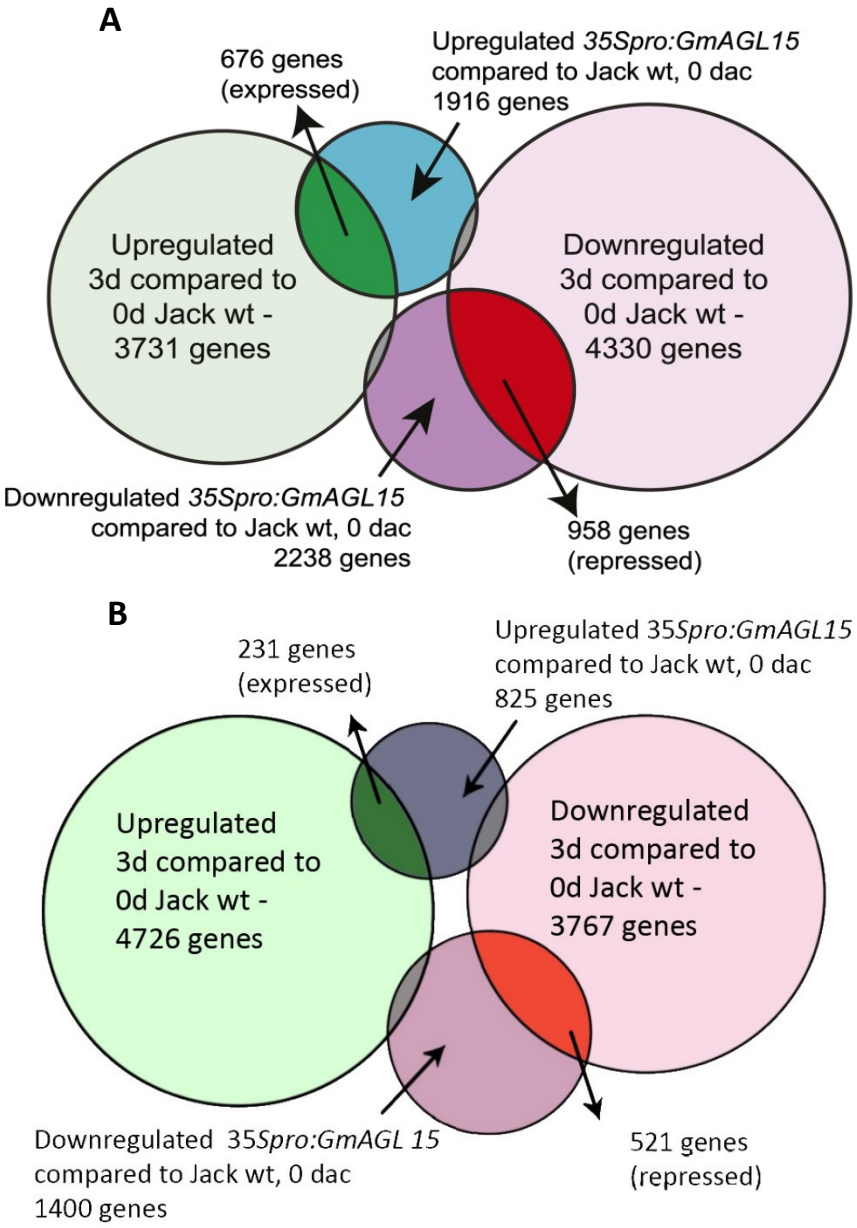



**Figure 5.2** Overlap between genes responsive to *35Spro:GmAGL15* compared to wild type and the genes of wild type that are responsive to 3 days on D40 medium compared with at 0 dac.

A. Results of previous microarray experiments from duplicates (Zheng and Perry 2014).
B. Results of RNA-seq experiments from triplicates in this study.

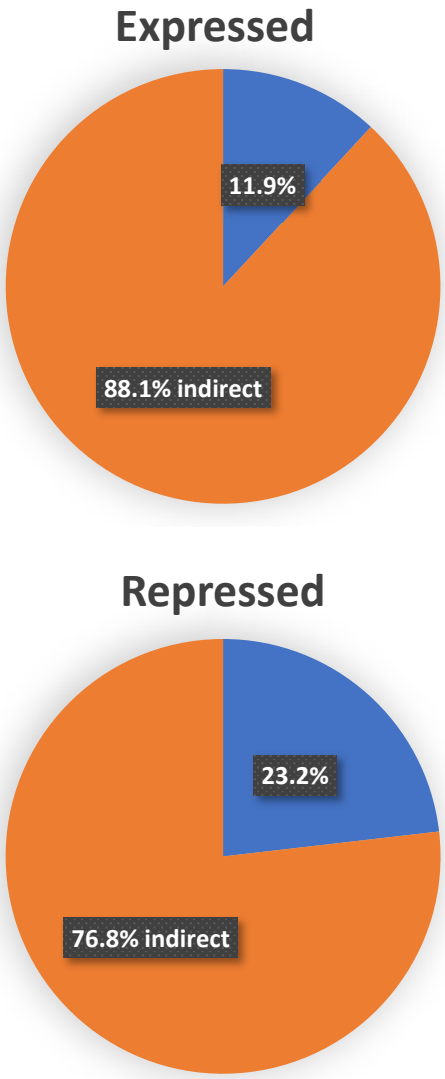



**Figure 5.3** Fraction of expressed and repressed genes that are regulated by AGL15 in Arabidopsis.

Orange: the indirect targets that respond to AGL15 accumulation but are not bound by AGL15.

Blue: the direct targets that respond to AGL15 accumulation and also are bound by AGL15.

**References**


Adamczyk, B. J., M. D. Lehti-Shiu and D. E. Fernandez (2007). "The MADS domain factors AGL15 and AGL18 act redundantly as repressors of the floral transition in Arabidopsis." Plant J **50**(6): 1007-1019.

Adhikari, P. B., X. Y. Liu, X. Y. Wu, S. W. Zhu and R. D. Kasahara (2020). "Fertilization in flowering plants: an odyssey of sperm cell delivery." Plant Molecular Biology **103**: 9-32.

Aida, M., D. Beis, R. Heidstra, V. Willemsen, I. Blilou, C. Galinha, L. Nussaume, Y. S. Noh, R. Amasino and B. Scheres (2004). "The PLETHORA genes mediate patterning of the Arabidopsis root stem cell niche." Cell **119**(1): 109-120.

Ashraf, M. and L. J. C. R. i. P. S. Wu (1994). "Breeding for salinity tolerance in plants." Critical Reviews in Plant Sciences **13**(1): 17-42.

Bargmann, B. O. R., A. Marshall-Colon, I. Efroni, S. Ruffel, K. D. Birnbaum, G. M. Coruzzi and G. Krouk (2013). "TARGET: A Transient Transformation System for Genome-Wide Transcription Factor Target Discovery." Molecular Plant **6**(3): 978-980.

Batista, R. A. and C. Kohler (2020). "Genomic imprinting in plants-revisiting existing models." Genes & Development **34**(1-2): 24-36.

Baud, S., Z. Kelemen, J. Thevenin, C. Boulard, S. Blanchet, A. To, M. Payre, N. Berger, D. Effroy-Cuzzi, J. M. Franco-Zorrilla, M. Godoy, R. Solano, E. Thevenon, F. Parcy, L. Lepiniec and B. Dubreucq (2016). "Deciphering the molecular mechanisms underpinning the transcriptional control of gene expression by master transcriptional regulators in Arabidopsis seed." Plant Physiol **171**(2): 1099-1112.

Baumlein, H., S. Misera, H. Luerssen, K. Kolle, C. Horstmann, U. Wobus and A. J. Muller (1994). "The *FUS3* gene of *Arabidopsis thaliana* is a regulator of gene expression during late embryogenesis."Plant J **6**(3): 379-387.

Bedi, S., S. Sengupta, A. Ray and R. N. Chaudhuri (2016). "ABI3 mediates dehydration stress recovery response in Arabidopsis thaliana by regulating expression of downstream genes." Plant Science **250**: 125-140.

Bensmihen, S., S. Rippa, G. Lambert, D. Jublot, V. Pautot, F. Granier, J. Giraudat and F. Parcy (2002). "The homologous ABI5 and EEL transcription factors function antagonistically to fine-tune gene expression during late embryogenesis." Plant Cell **14**(6): 1391-1403.

Bentsink, L., J. Jowett, C. J. Hanhart and M. Koornneef (2006). "Cloning of DOG1, a quantitative trait locus controlling seed dormancy in Arabidopsis." Proc Natl Acad Sci U S A **103**(45): 17042-17047.

Berger, F. (2003). "Endosperm: The crossroad of seed development." Current Opinion in Plant Biology **6**(1): 42-50.

Bleckmann, A., S. Alter and T. Dresselhaus (2014). "The beginning of a seed: regulatory mechanisms of double fertilization." Frontiers in Plant Science **5**: article no. 452.

Boisnard-Lorig, C., A. Colon-Carmona, M. Bauch, S. Hodge, P. Doerner, E. Bancharel, C. Dumas, J. Haseloff and F. Berger (2001). "Dynamic analyses of the expression of the HISTONE:: YFP fusion protein in Arabidopsis show that syncytial endosperm is divided in mitotic domains." Plant Cell **13**(3): 495-509.

Boulard, C., A. Fatihi, L. Lepiniec and B. Dubreucq (2017). "Regulation and evolution of the interaction of the seed B3 transcription factors with NF-Y subunits." Biochimica Et Biophysica Acta-Gene Regulatory Mechanisms **1860**(10): 1069-1078.

Boulard, C., J. Thevenin, O. Tranquet, V. Laporte, L. Lepiniec and B. Dubreucq (2018). "LEC1 (NF-YB9) directly interacts with LEC2 to control gene expression in seed." Biochimica Et Biophysica Acta-Gene Regulatory Mechanisms **1861**(5): 443-450.

Boutilier, K., R. Offringa, V. K. Sharma, H. Kieft, T. Ouellet, L. M. Zhang, J. Hattori, C. M. Liu, A. A. M. van Lammeren, B. L. A. Miki, J. B. M. Custers and M. M. V. Campagne (2002). "Ectopic expression of BABY BOOM triggers a conversion from vegetative to embryonic growth." Plant Cell **14**(8): 1737-1749.

Bozhkov, P. V., L. H. Filonova and M. F. Suarez (2005). "Programmed cell death in plant embryogenesis." Current Topics in Developmental Biology **67**: 135-179.

Bratzel, F., G. Lopez-Torrejon, M. Koch, J. C. Del Pozo and M. Calonje (2010). "Keeping cell identity in Arabidopsis requires PRC1 RING-finger homologs that catalyze H2A monoubiquitination." Current Biology **20**(20): 1853-1859.

Braybrook, S. A. and J. J. Harada (2008). "LECs go crazy in embryo development." Trends in Plant Science **13**(12): 624-630.

Braybrook, S. A., S. L. Stone, S. Park, A. Q. Bui, B. H. Le, R. L. Fischer, R. B. Goldberg and J. J. Harada (2006). "Genes directly regulated by LEAFY COTYLEDON2 provide insight into the control of embryo maturation and somatic embryogenesis." Proc Natl Acad Sci U S A **103**(9): 3468-3473.

Busch, W., A. Miotk, F. D. Ariel, Z. Zhao, J. Forner, G. Daum, T. Suzaki, C. Schuster, S. J. Schultheiss, A. Leibfried, S. Haubeiss, N. Ha, R. L. Chan and J. U. Lohmann (2010). "Transcriptional control of a plant stem cell niche." Developmental Cell **18**(5): 849-861.

Carbonero, P., R. Iglesias-Fernandez and J. Vicente-Carbajosa (2017). "The AFL subfamily of B3 transcription factors: evolution and function in angiosperm seeds." J Exp Bot **68**(4): 871-880.

Carmen, A. A., S. E. Rundlett and M. Grunstein (1996). "HDA1 and HDA3 are components of a yeast histone deacetylase (HDA) complex." Journal of Biological Chemistry **271**(26): 15837-15844.

Carrera, E., T. Holman, A. Medhurst, W. Peer, H. Schmuths, S. Footitt, F. L. Theodoulou and M. J. Holdsworth (2007). "Gene expression profiling reveals defined functions of the ATP-binding cassette transporter COMATOSE late in phase II of germination." Plant Physiol **143**(4): 1669-1679.

Casson, S., M. Spencer, K. Walker and K. Lindsey (2005). "Laser capture microdissection for the analysis of gene expression during embryogenesis of Arabidopsis." Plant J **42**(1): 111-123.

Chahtane, H., W. Kim and L. Lopez-Molina (2017). "Primary seed dormancy: a temporally multilayered riddle waiting to be unlocked." J Exp Bot **68**(4): 857-869.

Chandler, J., J. Nardmann and W. Werr (2008). "Plant development revolves around axes." Trends in Plant Science **13**(2): 78-84.

Chang, S., J. Puryear and J. Cairney (1993). "A simple and efficient method for isolating RNA from pine trees." Plant Molecular Biology Reporter **11**: 113-116.

Chanvivattana, Y., A. Bishopp, D. Schubert, C. Stock, Y. H. Moon, Z. R. Sung and J. Goodrich (2004). "Interaction of polycomb-group proteins controlling flowering in Arabidopsis." Development **131**(21): 5263-5276.

Chatr-Aryamontri A, R. Oughtred, L. Boucher, J. Rust, C. Chang, N. K. Kolas, L. O'Donnell, S. Oster, C. Theesfeld, A. Sellam, C. Stark, B. J. Breitkreutz, K. Dolinski and M. Tyers (2017). "The BioGRID interaction database: 2017 update." Nucleic Acids Res. **45**(D1): D369-D379.

Chaudhury, A. M., L. Ming, C. Miller, S. Craig, E. S. Dennis and W. J. Peacock (1997). "Fertilization-independent seed development in *Arabidopsis thaliana*." Proc Natl Acad Sci U S A **94**(8): 4223-4228.

Chen, D. H., A. Molitor, C. L. Liu and W. H. Shen (2010). "The Arabidopsis PRC1-like ring-finger proteins are necessary for repression of embryonic traits during vegetative growth." Cell Research **20**(12): 1332-1344.

Chen, N. C., V. Veerappan, H. Abdelmageed, M. Kang and R. D. Allen (2018). "HSI2/VAL1 silences *AGL15* to regulate the developmental transition from seed maturation to vegetative growth in Arabidopsis." Plant Cell **30**(3): 600-619.

Chen, N. C., H. Wang, H. Abdelmageed, V. Veerappan, M. Tadege and R. D. Allen (2020). "HSI2/VAL1 and HSL1/VAL2 function redundantly to repress DOG1 expression in Arabidopsis seeds and seedlings." New Phytologist **227**(3): 840-856

Chhun, T., S. Y. Chong, B. S. Park, E. C. C. Wong, J. L. Yin, M. Kim and N. H. Chua (2016). "HSI2 repressor recruits MED13 and HDA6 to down-regulate seed maturation gene expression directly during Arabidopsis early seedling growth." Plant and Cell Physiology **57**(8): 1689-1706.

Chiu, R. S., H. Nahal, N. J. Provart and S. Gazzarrini (2012). "The role of the Arabidopsis FUSCA3 transcription factor during inhibition of seed germination at high temperature." BMC Plant Biology **12**: article no. 15.

Chiu, R. S., Y. Saleh and S. Gazzarrini (2016). "Inhibition of FUSCA3 degradation at high temperature is dependent on ABA signaling and is regulated by the ABA/GA ratio." Plant Signaling & Behavior **11**(11): article no. e1247137.

Clough, S. J. and A. F. Bent (1998). "Floral dip: a simplified method for Agrobacterium-mediated transformation of Arabidopsis thaliana." Plant J **16**(6): 735-743.

Coen, O. and E. Magnani (2018). "Seed coat thickness in the evolution of angiosperms." Cellular and Molecular Life Sciences **75**(14): 2509-2518.

Colette, A., K.-J. Lu and D. Weijers (2015). "Building a plant: cell fate specification in the early Arabidopsis embryo." Development **142**(3): 420-430.

Conesa, A., P. Madrigal, S.Tarazona, D. Gomez-Cabrero, A. Cervera, A. McPherson, M.W. Szcześniak, D.J. Gaffney, L.L. Elo, X. Zhang, and A. Mortazavi (2016). "A survey of best practices for RNA-seq data analysis." Genome Biology **17**: article no. 13

Costa, M. C. D., K. Righetti, H. Nijveen, F. Yazdanpanah, W. Ligterink, J. Buitink and H. W. M. Hilhorst (2015). "A gene co-expression network predicts functional genes controlling the re-establishment of desiccation tolerance in germinated Arabidopsis thaliana seeds." Planta **242**(2): 435-449.

Cutler, S. and P. McCourt (2005). "Dude, where's my phenotype? Dealing with redundancy in signaling networks." Plant Physiol **138**(2): 558-559.

De Ruijter, A. J., A. H. Van Gennip, H. N. Caron, K. Stephan and A. B. Van Kuilenburg (2003). "Histone deacetylases (HDACs): characterization of the classical HDAC family." Biochemical Journal **370**(3): 737-749.

De Smet, I., S. Lau, U. Mayer and G. Jürgens (2010). "Embryogenesis–the humble beginnings of plant life." Plant J **61**(6): 959-970.

Delmas, F., S. Sankaranarayanan, S. Deb, E. Widdup, C. Bournonville, N. Bollier, J. G. B. Northey, P. McCourt and M. A. Samuel (2013). "ABI3 controls embryo degreening through Mendel's I locus." Proc Natl Acad Sci U S A **110**(40): E3888-E3894.

Dey, N., S. Sarkar, S. Acharya and I. Maiti (2015). "Synthetic promoters in planta." Planta **242**(5): 1077-1094.

El Ouakfaoui, S., J. Schnell, A. Abdeen, A. Colville, H. Labbe, S. Y. Han, B. Baum, S. Laberge and B. Miki (2010). "Control of somatic embryogenesis and embryo development by AP2 transcription factors." Plant Molecular Biology **74**(4-5): 313-326.

Elhiti, M., C. Stasolla and A. Wang (2013). "Molecular regulation of plant somatic embryogenesis." In Vitro Cellular & Developmental Biology-Plant **49**(6): 631-642.

Fatema, U., M. F. Ali, Z. Hu, A. J. Clark and T. Kawashima (2019). "Gamete nuclear migration in animals and plants." Frontiers in Plant Science **10**: article no. 517.

Faure, J. E., N. Rotman, P. Fortune and C. Dumas (2002). "Fertilization in Arabidopsis thaliana wild type: Developmental stages and time course." Plant J **30**(4): 481-488.

Feeney, M., L. Frigerio, Y. H. Cui and R. Menassa (2013). "Following vegetative to embryonic cellular changes in leaves of Arabidopsis overexpressing *LEAFY COTYLEDON2*." Plant Physiol **162**(4): 1881-1896.

Feher, A. (2019). "Callus, Dedifferentiation, Totipotency, Somatic Embryogenesis: What These Terms Mean in the Era of Molecular Plant Biology?" Frontiers in Plant Science **10**.

Feng, J. and L. Ma (2017). " A method for characterizing embryogenesis in Arabidopsis. ." Journal of Visualized Experiments : JoVE **126**: article no. 55969. https://doi.org/55910.53791/55969.

Fernandez, D. E., G. R. Heck, S. E. Perry, S. E. Patterson, A. B. Bleecker and S.-C. Fang (2000). "The embryo MADS domain factor AGL15 acts postembryonically: inhibition of perianth senescence and abscission via constitutive expression." Plant Cell **12**(2): 183-197.

Figueiredo, D. D. and C. Kohler (2016). "Bridging the generation gap: communication between maternal sporophyte, female gametophyte and fertilization products." Current Opinion in Plant Biology **29**: 16-20.

Finkelstein, R. (2013). "Abscisic acid synthesis and response." The Arabidopsis book/American Society of Plant Biologists **11**.

Florez, S. L., R. L. Erwin, S. N. Maximova, M. J. Guiltinan and W. R. Curtis (2015). "Enhanced somatic embryogenesis in Theobroma cacao using the homologous BABY BOOM transcription factor." Bmc Plant Biology **15**: 12.

Gallois, J. L., F. R. Nora, Y. Mizukami and R. Sablowski (2004). "WUSCHEL induces shoot stem cell activity and developmental plasticity in the root meristem." Genes & Development **18**(4): 375-380.

Gallois, J. L., C. Woodward, G. V. Reddy and R. Sablowski (2002). "Combined SHOOT MERISTEMLESS and WUSCHEL trigger ectopic organogenesis in Arabidopsis." Development **129**(13): 3207-3217.

Gazzarrini, S., Y. Tsuchiya, S. Lumba, M. Okamoto and P. McCourt (2004). "The transcription factor FUSCA3 controls developmental timing in Arabidopsis through the hormones gibberellin and abscisic acid." Developmental Cell **7**(3): 373-385.

Gepstein, S. and B. A. Horwitz (1995). "The impact of Arabidopsis research on plant biotechnology." Biotechnol Adv **13**(3): 403-414.

Ghelis, T., G. Bolbach, G. Clodic, Y. Habricot, E. Miginiac, B. Sotta and E. Jeannette (2008). "Protein tyrosine kinases and protein tyrosine phosphatases are involved in abscisic acid-dependent processes in Arabidopsis seeds and suspension cells." Plant Physiol **148**(3): 1668-1680.

Giraudat, J., B. M. Hauge, C. Valon, J. Smalle, F. Parcy and H. M. Goodman (1992). "Isolation of the Arabidopsis *ABI3* gene by positional cloning." Plant Cell **4**(10): 1251-1261.

Goldberg, R. B., G. De Paiva and R. Yadegari (1994). "Plant embryogenesis: zygote to seed." Science **266**(5185): 605-614.

Gonzalez-Morales, S. I., R. A. Chavez-Montes, C. Hayano-Kanashiro, G. Alejo-Jacuinde, T. Y. Rico-Cambron, S. de Folter and L. Herrera-Estrella (2016). "Regulatory network analysis reveals novel regulators of seed desiccation tolerance in Arabidopsis thaliana." Proc Natl Acad Sci U S A **113**(35): E5232-E5241.

Guo, F. D., C. L. Liu, H. Xia, Y. P. Bi, C. Z. Zhao, S. Z. Zhao, L. Hou, F. G. Li and X. J. Wang (2013). "Induced expression of *AtLEC1* and *AtLEC2* differentially promotes somatic embryogenesis in transgenic tobacco plants." PLoS One **8**(8): article no. e71714.

Gutierrez, L., O. V. Wuytswinkel, M. Castelain and C. Bellini (2007). "Combined networks regulating seed maturation." Trends in Plant Science **127**(7): 294-300.

Han, J. D., X. Li, C. K. Jiang, G. K. S. Wong, C. J. Rothfels and G. Y. Rao (2017). "Evolutionary analysis of the LAFL genes involved in the land plant seed maturation program." Frontiers in Plant Science **8**: article no. 439.

Han, Q., A. Bartels, X. Cheng, A. Meyer, Y. Q. C. An, T. F. Hsieh and W. Y. Xiao (2019). "Epigenetics regulates reproductive development in plants." Plants-Basel **8**(12): article no. 564.

Han, X., L. J. Huang, D. Feng, W. H. Jiang, W. Z. Miu and N. Li (2019). "Plasmodesmata-related structural and functional proteins: The long sought-after secrets of a cytoplasmic channel in plant cell walls." International Journal of Molecular Sciences **20**(12): article no. 2946.

Haque, M. E., B. Han, B. Wang, Y. Wang and A. Z. Liu (2018). "Development of an efficient chromatin immunoprecipitation method to investigate protein-DNA interaction in oleaginous castor bean seeds." Plos One **13**(5).

Harada, J. J. (2001). "Role of Arabidopsis LEAFY COTYLEDON genes in seed development." Journal of Plant Physiol **158**(4): 405-409.

Harb, A., A. Krishnan, M. M. R. Ambavaram and A. Pereira (2010). "Molecular and Physiological Analysis of Drought Stress in Arabidopsis Reveals Early Responses Leading to Acclimation in Plant Growth." Plant Physiol **154**(3): 1254-1271.

Harding, E. W., W. Tang, K. W. Nichols, D. E. Fernandez and S. E. Perry (2003). "Expression and maintenance of embryogenic potential is enhanced through constitutive expression of *AGAMOUS-Like 15*." Plant Physiol. **133**: 653-663.

Hecht, V., J. P. Vielle-Calzada, M. V. Hartog, E. D. L. Schmidt, K. Boutilier, U. Grossniklaus and S. C. de Vries (2001). "The Arabidopsis *SOMATIC EMBRYOGENESIS RECEPTOR KINASE 1* gene is expressed in developing ovules and embryos and enhances embryogenic competence in culture." Plant Physiol **127**(3): 803-816.

Heck, G. R., S. E. Perry, K. W. Nichols and D. E. Fernandez (1995). "AGL15, a MADS domain protein expressed in developing embryos." Plant Cell **7**(8): 1271-1282.

Hehenberger, E., D. Kradolfer and C. Köhler (2012). "Endosperm cellularization defines an important developmental transition for embryo development." Development **139**(11): 2031-2039.

Hill, K., H. Wang and S. E. Perry (2008). "A transcriptional repression motif in the MADS factor AGL15 is involved in recruitment of histone deacetylase complex components." Plant J **53**(1):172-185.

Holdsworth, M. J., L. Bentsink and W. J. J. Soppe (2008). "Molecular networks regulating Arabidopsis seed maturation, after‐ripening, dormancy and germination." New Phytologist **179**: 33-54.

Horstman, A., M. F. Li, I. Heidmann, M. Weemen, B. J. Chen, J. M. Muino, G. C. Angenent and K. Boutilier (2017). "The BABY BOOM transcription factor activates the LEC1-ABI3-FUS3-LEC2 network to induce somatic embryogenesis." Plant Physiol **175**(2): 848-857.

Huang, M. K., Y. L. Hu, X. Liu, Y. G. Li and X. L. Hou (2015). "Arabidopsis LEAFY COTYLEDON1 mediates postembryonic development via interacting with PHYTOCHROME-INTERACTING FACTOR4." Plant Cell **27**(11): 3099-3111.

Huang, X. Z., X. Y. Zhang, Z. Z. Gong, S. H. Yang and Y. T. Shi (2017). "ABI4 represses the expression of type-A ARRs to inhibit seed germination in Arabidopsis." Plant J **89**(2): 354-365.

Hundertmark, M. and D. K. Hincha (2008). "LEA (late embryogenesis abundant) proteins and their encoding genes in Arabidopsis thaliana." BMC Genomics **9**: 118.

Ikeuchi, M., A. Iwase, B. Rymen, H. Harashima, M. Shibata, M. Ohnuma, C. Breuer, A. K. Morao, M. de Lucas, L. De Veylder, J. Goodrich, S. M. Brady, F. Roudier and K. Sugimoto (2015). "PRC2 represses dedifferentiation of mature somatic cells in Arabidopsis." Nature Plants **1**(7): 1-7.

Jacob, D. and J. Brian (2020). "The short and intricate life of the suspensor." Physiologia Plantarum **169**: 110-121.

Jakoby, M., B. Weisshaar, W. Droge-Laser, J. Vicente-Carbajosa, J. Tiedemann, T. Kroj, F. Parcy and bZIP Research Group (2002). "bZIP transcription factors in Arabidopsis." Trends in Plant Science **7**(3): 106-111.

Jha, P., S. J. Ochatt and V. Kumar (2020). "WUSCHEL: a master regulator in plant growth signaling." Plant Cell Reports **39**(4): 431-444.

Ji, H. K., H. Jiang, W. X. Ma, D. S. Johnson, R. M. Myers and W. H. Wong (2008). "An integrated software system for analyzing ChIP-chip and ChIP-seq data." Nature Biotechnology **26**(11): 1293-1300.

Jia, H. Y., D. R. McCarty and M. Suzuki (2013). "Distinct Roles of LAFL Network Genes in Promoting the Embryonic Seedling Fate in the Absence of VAL Repression." Plant Physiol **163**(3): 1293-1305.

Jo, L., J. M. Pelletier and J. J. Harada (2019). "Central role of the LEAFY COTYLEDON1 transcription factor in seed development." Journal of Integrative Plant Biology **61**(5): 564-580.

Junker, A. and H. Baumlein (2012). "Multifunctionality of the LEC1 transcription factor during plant development." Plant Signaling & Behavior **7**(12): 1718-1720.

Junker, A., G. Monke, T. Rutten, J. Keilwagen, M. Seifert, M. N. T. Tuyet, J. P. Renou, S. Balzergue, P. Viehover, U. Hahnel, J. Ludwig-Muller, L. Altschmied, U. Conrad, B. Weisshaar and H. Baumlein (2012). "Elongation-related functions of LEAFY COTYLEDON1 during the development of Arabidopsis thaliana." Plant J **71**(3): 427-442.

Jürgens, G. (2001). "Apical–basal pattern formation in Arabidopsis embryogenesis." The EMBO journal **20**(14): 3609-3616.

Kagale, S. and K. Rozwadowski (2011). "EAR motif-mediated transcriptional repression in plants: an underlying mechanism for epigenetic regulation of gene expression." Epigenetics **6**(2): 141-146.

Kagaya, Y., R. Okuda, A. Ban, R. Toyoshima, K. Tsutsumida, H. Usui, A. Yamamoto and T. Hattori (2005). "Indirect ABA-dependent regulation of seed storage protein genes by FUSCA3 transcription factor in Arabidopsis." Plant and Cell Physiology **46**(2): 300-311.

Karlova, R., S. Boeren, E. Russinova, J. Aker, J. Vervoort and S. de Vries (2006). "The Arabidopsis SOMATIC EMBRYOGENESIS RECEPTOR-LIKE KINASE1 protein complex includes BRASSINOSTEROID-INSENSITIVE1." Plant Cell **18**(3): 626-638.

Kasim, W. A., M. E. Osman, M. N. Omar, I. A. A. El-Daim, S. Bejai and J. Meijer (2013). "Control of drought stress in wheat using plant-growth-promoting bacteria." Journal of plant growth regulation **32**(1): 122-130.

Kawagoe, Y. and N. Murai (1996). "A novel basic region/helix-loop-helix protein binds to a G-box motif CACGTG of the bean seed storage protein beta-phaseolin gene." Plant Science **116**(1): 47-57.

Keith, K., M. Kraml, N. G. Dengler and P. McCourt (1994). "FUSCA3 - A heterochronic mutation affecting late embryo development in Arabidopsis." Plant Cell **6**(5): 589-600.

Kimata, Y., T. Higaki, T. Kawashima, D. Kurihara, Y. Sato, T. Yamada, S. Hasezawa, F. Berger, T. Higashiyama and M. Ueda (2016). "Cytoskeleton dynamics control the first asymmetric cell division in Arabidopsis zygote." Proc Natl Acad Sci U S A **113**(49): 14157-14162.

Klepikova, A. V., A. S. Kasianov, E. S. Gerasimov, M. D. Logacheva and A. A. Penin (2016). "A high resolution map of the *Arabidopsis thaliana* developmental transcriptome based on RNA-seq profiling." Plant J **88**(6): 1058-1070.

Koszegi, D., A. J. Johnston, T. Rutten, A. Czihal, L. Altschmied, J. Kumlehn, S. E. J. Wust, O. Kirioukhova, J. Gheyselinck, U. Grossniklaus and H. Baumlein (2011). "Members of

the RKD transcription factor family induce an egg cell-like gene expression program." Plant J **67**(2): 280-291.
Kotak, S., E. Vierling, H. Baumlein and P. von Koskull-Doring (2007). "A novel transcriptional cascade regulating expression of heat stress proteins during seed development of Arabidopsis." Plant Cell **19**(1): 182-195.
Kozomara, A., M. Birgaoanu and S. Griffiths-Jones (2019). "miRBase: from microRNA sequences to function." Nucl. Acids Res. **47**: D155-D162.
Kumlehn, J., H. Lorz and E. Kranz (1998). "Differentiation of isolated wheat zygotes into embryos and normal plants." Planta **205**(3): 327-333.
Kwong, R. W., A. Q. Bui, H. Lee, L. W. Kwong, R. L. Fischer, R. B. Goldberg and J. J. Harada (2003). "LEAFY COTYLEDON1-LIKE defines a class of regulators essential for embryo development." Plant Cell **15**(1): 5-18.
Lackman, P., M. González-Guzmán, S. Tilleman, I. Carqueijeiro, A. C. Pérez, T. Moses, M. Seo, Y. Kanno, S. T. Häkkinen and M. C. Van Montagu (2011). "Jasmonate signaling involves the abscisic acid receptor PYL4 to regulate metabolic reprogramming in Arabidopsis and tobacco." Proc Natl Acad Sci U S A **108**(14): 5891-5896.
Lafon-Placette, C. and C. Köhler (2014). "Embryo and endosperm, partners in seed development." Current Opinion in Plant Biology **17**: 64-69.
Lau, S., D. Slane, O. Herud, J. Kong and G. Jürgens (2012). "Early embryogenesis in flowering plants: setting up the basic body pattern." Annual Review of Plant Biology **63**: 483-506.
Laux, T. and G. Jurgens (1997). "Embryogenesis: A new start in life." Plant Cell **9**(7): 989-1000.
Laux, T., T. Würschum and H. Breuninger (2004). "Genetic regulation of embryonic pattern formation." Plant Cell **16**: S190-S202.
Le, B. H., C. Cheng, A. Q. Bui, J. A. Wagmaister, K. F. Henry, J. Pelletier, L. Kwong, M. Belmonte, R. Kirkbride and S. Horvath (2010). "Global analysis of gene activity during Arabidopsis seed development and identification of seed-specific transcription factors." Proc Natl Acad Sci U S A **107**(18): 8063-8070.
Lee, H. S., R. L. Fischer, R. B. Goldberg and J. J. Harada (2003). "Arabidopsis LEAFY COTYLEDON1 represents a functionally specialized subunit of the CCAAT binding transcription factor." Proc Natl Acad Sci U S A **100**(4): 2152-2156.
Lee, J., K. He, V. Stolc, H. Lee, P. Figueroa, Y. Gao, W. Tongprasit, H. Zhao, I. Lee and X. W. Deng (2007). "Analysis of transcription factor HY5 genomic binding sites revealed its hierarchical role in light regulation of development." Plant Cell **19**: 731-749.
Leprince, O., A. Pellizzaro, S. Berriri and J. Buitink (2017). "Late seed maturation: drying without dying." J Exp Bot **68**(4): 827-841.
Li, J. and F. Berger (2012). "Endosperm: food for humankind and fodder for scientific discoveries." New Phytologist **195**(2): 290-305.
Li, J., G. Brader and E. T. Palva (2004). "The WRKY70 transcription factor: a node of convergence for jasmonate-mediated and salicylate-mediated signals in plant defense." Plant Cell **16**(2): 319-331.
Lim, S., J. Park, N. Lee, J. Jeong, S. Toh, A. Watanabe, J. Kim, H. Kang, D. H. Kim, N. Kawakami and G. Choi (2013). "ABA-INSENSITIVE3, ABA-INSENSITIVE5, and DELLAs

Interact to Activate the Expression of SOMNUS and Other High-Temperature-Inducible Genes in Imbibed Seeds in Arabidopsis." Plant Cell **25**(12): 4863-4878.

Liu, W. w., J. Meng, J. Cui and Y.-s. Luan (2017). "Characterization and Function of MicroRNA*s in Plants." Frontiers in Plant Science **8**: 2200.

Liu, X. D., H. Zhang, Y. Zhao, Z. Y. Feng, Q. Li, H. Q. Yang, S. Luan, J. M. Li and Z. H. He (2013). "Auxin controls seed dormancy through stimulation of abscisic acid signaling by inducing ARF-mediated ABI3 activation in Arabidopsis." Proc Natl Acad Sci U S A **110**(38): 15485-15490.

Liu, Z., X. X. Ge, W. M. Qiu, J. M. Long, H. H. Jia, W. Yang, M. Dutt, X. M. Wu and W. W. Guo (2018). "Overexpression of the CsFUS3 gene encoding a B3 transcription factor promotes somatic embryogenesis in Citrus." Plant Science **277**: 121-131.

Lopez-Molina, L., S. Mongrand, N. Kinoshita, and N.-H. J. G. Chua (2003). "AFP is a novel negative regulator of ABA signaling that promotes ABI5 protein degradation." Development **17**(3): 410-418.

López-Ribera, I., J. L. La Paz, C. Repiso, N. García, M. Miquel, M. L. Hernández, J. M. Martínez-Rivas and C. M. Vicient (2014). "The evolutionary conserved oil body associated protein OBAP1 participates in the regulation of oil body size." Plant Physiol **164**(3): 1237-1249.

Lotan, T., M. Ohto, K. M. Yee, M. A. L. West, R. Lo, R. W. Kwong, K. Yamagishi, R. L. Fischer, R. B. Goldberg and J. J. Harada (1998). "Arabidopsis LEAFY COTYLEDON1 is sufficient to induce embryo development in vegetative cells." Cell **93**(7): 1195-1205.

Lu, Q. S., J. dela Paz, A. Pathmanathan, R. S. Chiu, A. Y. L. Tsai and S. Gazzarrini (2010). "The C-terminal domain of FUSCA3 negatively regulates mRNA and protein levels, and mediates sensitivity to the hormones abscisic acid and gibberellic acid in Arabidopsis." Plant J **64**(1): 100-113.

Luerssen, K., V. Kirik, P. Herrmann and S. Misera (1998). "FUSCA3 encodes a protein with a conserved VP1/ABI3-like B3 domain which is of functional importance for the regulation of seed maturation in *Arabidopsis thaliana*." Plant J **15**(6): 755-764.

Lumba, S., Y. Tsuchiya, F. Delmas, J. Hezky, N. J. Provart, Q. Shi Lu, P. McCourt and S. Gazzarrini (2012). "The embryonic leaf identity gene FUSCA3 regulates vegetative phase transitions by negatively modulating ethylene-regulated gene expression in Arabidopsis." BMC biology **10**: 8.

Lynch, T. J., B. J. Erickson, D. R. Miller and R. R. Finkelstein (2017). "ABI5-binding proteins (AFPs) alter transcription of ABA-induced genes via a variety of interactions with chromatin modifiers." Plant Molecular Biology **93**(4-5): 403-418.

Manfre, A. J., L. M. Lanni and W. R. Marcotte (2006). "The Arabidopsis group 1 LATE EMBRYOGENESIS ABUNDANT protein ATEM6 is required for normal seed development." Plant Physiol **140**(1): 140-149.

Mansfield, S. and L. Briarty (1991). "Early embryogenesis in *Arabidopsis thaliana*. II. The developing embryo." Canadian Journal of Botany **69**(3): 461-476.

Meinke, D. W. (1992). "A homeotic mutant of *Arabidopsis thaliana* with leafy cotyledons." Science **258**(5088): 1647-1650.

Meinke, D. W. (2020). "Genome-wide identification of EMBRYO-DEFECTIVE (EMB) genes required for growth and development in Arabidopsis." New Phytologist **226**(2): 306-325.

Meinke, D. W., L. H. Franzmann, T. C. Nickle and E. C. Yeung (1994). "Leafy cotyledon mutants of Arabidopsis." Plant Cell **6**(8): 1049-1064.

Mi, H., S. Poudel, A. Muruganujan, J. T. Casagrande and P. D. J. Thomas (2016). "PANTHER version 10: expanded protein families and functions, and analysis tools." Nucleic Acids Research **44**(D1): D336-D342.

Monke, G., L. Altschmied, A. Tewes, W. Reidt, H. P. Mock, H. Baumlein and U. Conrad (2004). "Seed-specific transcription factors ABI3 and FUS3: molecular interaction with DNA." Planta **219**(1): 158-166.

Monke, G., M. Seifert, J. Keilwagen, M. Mohr, I. Grosse, U. Hähnel, A. Junker, B. Weisshaar, U. Conrad, H. Bäumlein and L. Altschmied (2012). "Toward the identification and regulation of the Arabidopsis thaliana ABI3 regulon." Nucleic Acids Research **40**: 8240-8254

Morcillo, F., A. Gallard, M. Pillot, S. Jouannic, F. Aberlenc-Bertossi, M. Collin, J. L. Verdeil and J. W. Tregear (2007). "EgAP2-1, an AINTEGUMENTA-like (AIL) gene expressed in meristematic and proliferating tissues of embryos in oil palm." Planta **226**(6): 1353-1362.

Mursyanti, E., P. Aziz, S. Moeljopawiro and E. Semiarti (2015). "Induction of somatic embryogenesis through overexpression of ATRKD4 genes in Phalaenopsis "Sogo Vivien"." Indonesian Journal of Biotechnology **20**(1): 42-53.

Nakabayashi, K., M. Okamoto, T. Koshiba, Y. Kamiya and E. Nambara (2005). "Genome-wide profiling of stored mRNA in Arabidopsis thaliana seed germination: epigenetic and genetic regulation of transcription in seed." Plant J **41**(5): 697-709.

Nakagawa, T., S. Ishiguro and T. Kimura (2009). "Gateway vectors for plant transformation." Plant Biotechnology **26**(3): 275-284.

Nakamura, S., T. J. Lynch and R. R. Finkelstein (2001). "Physical interactions between ABA response loci of Arabidopsis." Plant J **26**(6): 627-635.

Nakashima, K. and K. Yamaguchi-Shinozaki (2013). "ABA signaling in stress-response and seed development." Plant Cell Reports **32**(7): 959-970.

Nambara, E., S. Naito and P. McCourt (1992). "A mutant of Arabidopsis which is defective in seed development and storage protein accumulation is a new abi3 allele " Plant J **2**(435-441).

Nguyen, T.-P., G. Cueff, D. D. Hegedus, L. Rajjou and L. Bentsink (2015). "A role for seed storage proteins in Arabidopsis seed longevity. ." J. Exp. Bot. **66**(20): 6399-6413.

Nodine, M. D. and D. P. Bartel (2010). "MicroRNAs prevent precocious gene expression and enable pattern formation during plant embryogenesis." Genes & Development **24**(23): 2678-2692.

Nowack, M. K., P. E. Grini, M. J. Jakoby, M. Lafos, C. Koncz and A. Schnittger (2006). "A positive signal from the fertilization of the egg cell sets off endosperm proliferation in angiosperm embryogenesis." Nature Genetics **38**(1): 63-67.

Ogas, J., J.-C. Cheng, Z. R. Sung and C. Somerville (1997). "Cellular differentiation regulated by gibberellin in the *Arabidopsis thaliana pickle* mutant." Science **277**: 91-94.

Ogas, J., S. Kaufmann, J. Henderson and C. Somerville (1999). "PICKLE is a CHD3 chromatin-remodeling factor that regulates the transition from embryonic to vegetative development in Arabidopsis." Proc Natl Acad Sci U S A **96**(24): 13839-13844.

Oh, E., H. Kang, S. Yamaguchi, J. Park, D. Lee, Y. Kamiya and G. Choi (2009). "Genome-Wide Analysis of Genes Targeted by PHYTOCHROME INTERACTING FACTOR 3-LIKE5 during Seed Germination in Arabidopsis." Plant Cell **21**(2): 403-419.

Ohta, M., K. Matsui, K. Hiratsu, H. Shinshi and M. Ohme-Takagi (2001). "Repression domains of class II ERF transcriptional repressors share an essential motif for active repression." Plant Cell **13**: 1959-1968.

Olsen, O.-A. (2001). "Endosperm development: Cellularization and cell fate specification." Annual Review of Plant Physiol and Plant Molecular Biology **52**(1): 233-267.

Olsen, O.-A. (2004). "Nuclear endosperm development in cereals and *Arabidopsis thaliana*." Plant Cell **16**(suppl 1): S214-S227.

Otero, S., Y. Helariutta and Y. Benitez-Alfonso (2016). "Symplastic communication in organ formation and tissue patterning." Current Opinion in Plant Biology **29**: 21-28.

Palovaara, J., T. de Zeeuw and D. Weijers (2016). Tissue and organ initiation in the plant embryo: A first time for everything. Annual Review of Cell and Developmental Biology, Vol 32. R. Schekman. **32:** 47-75.

Parcy, F. and J. Giraudat (1997). "Interactions between the ABI1 and the ectopically expressed ABI3 genes in controlling abscisic acid responses in Arabidopsis vegetative tissues." Plant J **11**(4): 693-702.

Parcy, F., C. Valon, M. Raynal, P. Gaubier-Comella, M. Delseny and J. Giraudat (1994). "Regulation of gene expression programs during Arabidopsis seed development: roles of the ABI3 locus and of endogenous abscisic acid." Plant Cell **6**(11): 1567-1582.

Park, J., N. Lee, W. Kim, S. Lim and G. Choi (2011). "ABI3 and PIL5 Collaboratively Activate the Expression of SOMNUS by Directly Binding to Its Promoter in Imbibed Arabidopsis Seeds." Plant Cell **23**(4): 1404-1415.

Patharkar, O. R., T. A. Macken and J. C. Walker (2016). "Serine 231 and 257 of AGAMOUS-Like 15 are phosphorylated in floral receptacles." Plant Signaling & Behavior **11**(7): article no. UNSP e1199314.

Patharkar, O. R. and J. C. Walker (2015). "Floral organ abscission is regulated by a positive feedback loop." Proc Natl Acad Sci U S A **112**(9): 2906-2911.

Pelletier, J. M., R. W. Kwong, S. Park, B. H. Le, R. Baden, A. Cagliaria, M. Hashimoto, M. D. Munoz, R. L. Fischer, R. B. Goldberg and J. J. Harada (2017). "LEC1 sequentially regulates the transcription of genes involved in diverse developmental processes during seed development." Proc Natl Acad Sci U S A **114**(32): E6710-E6719.

Perez-Pascual, D., D. Jimenez-Guillen, H. Villanueva-Alonzo, R. Souza-Perera, G. Godoy-Hernandez and J. J. Zuniga-Aguilar (2018). "Ectopic expression of the Coffea canephora SERK1 homolog-induced differential transcription of genes involved in auxin metabolism and in the developmental control of embryogenesis." Physiologia Plantarum **163**(4): 530-551.

Peters, S., A. Egert, B. Stieger and F. Keller (2010). "Functional identification of Arabidopsis ATSIP2 (At3g57520) as an alkaline α -galactosidase with a substrate specificity for raffinose and an apparent sink-specific expression pattern." Plant and Cell Physiology **51**(10): 1815-1819.

Pignocchi, C., G. E. Minns, N. Nesi, R. Koumproglou, G. Kitsios, C. Benning, C. W. Lloyd, J. H. Doonan and M. J. Hills (2009). "ENDOSPERM DEFECTIVE1 is a novel microtubule-associated protein essential for seed development in Arabidopsis." Plant Cell **21**(1): 90-105.

Pizzio, G. A., L. Rodriguez, R. Antoni, M. Gonzalez-Guzman, C. Yunta, E. Merilo, H. Kollist, A. Albert and P. L. Rodriguez (2013). "The PYL4 A194T mutant uncovers a key role of PYR1-LIKE4/PROTEIN PHOSPHATASE 2CA interaction for abscisic acid signaling and plant drought resistance." Plant Physiol **163**(1): 441-455.

Rashid, S. Z., N. Yamaji and M. Kyo (2007). "Shoot formation from root tip region: a developmental alteration by WUS in transgenic tobacco." Plant Cell Reports **26**(9): 1449-1455.

Reeves, R. G. (1935). "Development of the cotton embryo." J. Agricultural Res. **51**: 935-944.

Rider, S. D., J. T. Henderson, R. E. Jerome, H. J. Edenberg, J. Romero-Severson and J. Ogas (2003). "Coordinate repression of regulators of embryonic identity by PICKLE during germination in Arabidopsis." Plant J **35**(1): 33-43.

Robert, H. S. (2019). "Molecular communication for coordinated seed and fruit development: What can we learn from auxin and sugars?" International Journal of Molecular Sciences **20**(4): article no. 936.

Roberts, E. H. (1973). "Predicting the storage life of seeds." Seed Science and Technology **1**: 499-514.

Robinson, S. J. and I. A. Parkin (2009). Bridging the gene-to-function knowledge gap through functional genomics. Plant Genomics, Springer**:** 153-173.

Robinson, S. J., L. H. Tang, B. A. Mooney, S. J. McKay, W. E. Clarke, M. G. Links, S. Karcz, S. Regan, Y.-Y. Wu and M. Y. Gruber (2009). "An archived activation tagged population of Arabidopsis thalianato facilitate forward genetics approaches." BMC Plant Biology **9**(1): 101.

Rodrigues, A. S. and C. M. Miguel (2017). "The pivotal role of small non-coding RNAs in the regulation of seed development." Plant Cell Reports **36**(5): 653-667.

Roscoe, T. J., V. Vaissayre, G. Paszkiewicz, F. Clavijo, Z. Kelemen, C. Michaud, L. Lepiniec, B. Dubreucq, D. X. Zhou and M. Devic (2019). "Regulation of FUSCA3 Expression During Seed Development in Arabidopsis." Plant and Cell Physiology **60**(2): 476-487.

Rounsley, S. D., G. S. Ditta and M. F. Yanofsky (1995). "Diverse roles for MADS box gene in Arabidopsis development." Plant Cell **7**(8): 1259-1269.

Ruduś, I., E. Kępczyńska and J. Kępczyński (2006). "Comparative efficacy of abscisic acid and methyl jasmonate for indirect somatic embryogenesis in Medicago sativa L." Plant Growth Regulation **48**(1): 1.

Ruduś, I., E. W. Weiler and E. Kępczyńska (2009). "Do stress-related phytohormones, abscisic acid and jasmonic acid play a role in the regulation of Medicago sativa L. somatic embryogenesis?" Plant Growth Regulation **59**(2): 159-169.

Sah, S. K., K. R. Reddy and J. Li (2016). "Abscisic acid and abiotic stress tolerance in crop plants." Frontiers in Plant Science **7**: 571.

Sano, N., L. Rajjou, H. M. North, I. Debeaujon, A. Marion-Poll and M. Seo (2016). "Staying Alive: Molecular Aspects of Seed Longevity." Plant and Cell Physiology **57**(4): 660-674.

Sasnauskas, G., E. Manakova, K. Lapenas, K. Kauneckaite and V. Siksnys (2018). "DNA recognition by Arabidopsis transcription factors ABI3 and NGA1." FEBS Journal **285**(21): 4041-4059.

Sattler, S. E., L. U. Gilliland, M. Magallanes-Lundback, M. Pollard and D. DellaPenna (2004). "Vitamin E is essential for seed longevity and for preventing lipid peroxidation during germination." Plant Cell **16**: 1419-1432.

Schmid, M., T. S. Davison, S. R. Henz, U. J. Pape, M. Demar, M. Vingron, B. Scholkopf, D. Weigel and J. U. Lohmann (2005). "A gene expression map of Arabidopsis thaliana development." Nature Genetics **37**(5): 501-506.

Schneider, A., D. Aghamirzaie, H. Elmarakeby, A. N. Poudel, A. J. Koo, L. S. Heath, R. Grene and E. Collakova (2016). "Potential targets of VIVIPAROUS1/ABI3-LIKE1 (VAL1) repression in developing Arabidopsis thaliana embryos." Plant J **85**(2): 305-319.

Schwartz, B. W., E. C. Yeung and D. W. Meinke (1994). "Disruption of morphogenesis and transformation of the suspensor in abnormal suspensor mutants of Arabidopsis." Development **120**(11): 3235-3245.

Serivichyaswat, P., H. S. Ryu, W. Kim, S. Kim, K. S. Chung, J. J. Kim and J. H. Ahn (2015). "Expression of the floral repressor miRNA156 is positively regulated by the AGAMOUS-like proteins AGL15 and AGL18." Molecules and Cells **38**(3): 259-266.

Setiari, N., A. Purwantoro, S. Moeljopawiro and E. Semiarti (2018). "Micropropagation of Dendrobium phalaenopsis Orchid Through Overexpression of Embryo Gene AtRKD4." Agrivita **40**(2): 284-294.

Shanker, A. K., M. Maheswari, S. Yadav, S. Desai, D. Bhanu, N. B. Attal and B. Venkateswarlu (2014). "Drought stress responses in crops." Functional integrative genomics **14**(1): 11-22.

Shu, J., C. Chen, R. K. Thapa, S. M. Bian, V. Nguyen, K. F. Yu, Z. C. Yuan, J. Liu, S. E. Kohalmi, C. L. Li and Y. H. Cui (2019). "Genome-wide occupancy of histone H3K27 methyltransferases CURLY LEAF and SWINGER in Arabidopsis seedlings." Plant Direct **3**(1): article no. Unsp 100.

Sivaramakrishna, D. (1978). "Size relationships of apical cell and basal cell in two-celled embryos in angiosperms." Canadian Journal of Botany **56**(12): 1434-1438.

Skubacz, A., A. Daszkowska-Golec and I. Szarejko (2016). "The role and regulation of ABI5 (ABA-Insensitive 5) in plant development, abiotic stress responses and phytohormone crosstalk." Frontiers in Plant Science **7**: 1884.

Sokol, A., A. Kwiatkowska, A. Jerzmanowski and M. Prymakowska-Bosak (2007). "Up-regulation of stress-inducible genes in tobacco and Arabidopsis cells in response

to abiotic stresses and ABA treatment correlates with dynamic changes in histone H3 and H4 modifications." Planta **227**(1): 245-254.

Song, C.-P. and D. W. Galbraith (2006). "AtSAP18, an orthologue of human SAP18, is involved in the regulation of salt stress and mediates transcriptional repression in Arabidopsis." Plant Molecular Biology **60**(2): 241-257.

Song, L., S. S. C. Huang, A. Wise, R. Castanon, J. R. Nery, H. M. Chen, M. Watanabe, J. Thomas, Z. Bar-Joseph and J. R. Ecker (2016). "A transcription factor hierarchy defines an environmental stress response network." Science **354**(6312).

Souter, M. and K. Lindsey (2000). "Polarity and signalling in plant embryogenesis." J Exp Bot **51**(347): 971-983.

Sprunck, S. (2020). "Twice the fun, double the trouble: gamete interactions in flowering plants." Current Opinion in Plant Biology **53**: 106-116.

Srinivasan, C., Z. R. Liu, I. Heidmann, E. D. J. Supena, H. Fukuoka, R. Joosen, J. Lambalk, G. Angenent, R. Scorza, J. B. M. Custers and K. Boutilier (2007). "Heterologous expression of the BABY BOOM AP2/ERF transcription factor enhances the regeneration capacity of tobacco (Nicotiana tabacum L.)." Planta **225**(2): 341-351.

Stone, S. L., S. A. Braybrook, S. L. Paula, L. W. Kwong, J. Meuser, J. Pelletier, T. F. Hsieh, R. L. Fischer, R. B. Goldberg and J. J. Harada (2008). "Arabidopsis LEAFY COTYLEDON2 induces maturation traits and auxin activity: Implications for somatic embryogenesis." Proc Natl Acad Sci U S A **105**(8): 3151-3156.

Stone, S. L., L. W. Kwong, K. M. Yee, J. Pelletier, L. Lepiniec, R. L. Fischer, R. B. Goldberg and J. J. Harada (2001). "LEAFY COTYLEDON2 encodes a B3 domain transcription factor that induces embryo development." Proc Natl Acad Sci U S A **98**(20): 11806-11811.

Strino, F. and M. J. B. b. Lappe (2016). "Identifying peaks in*-seq data using shape information." BMC bioinformatics **17**(5): S206.

Su, Y. H., X. Y. Zhao, Y. B. Liu, C. L. Zhang, S. D. O'Neill and X. S. Zhang (2009). "Auxin-induced WUS expression is essential for embryonic stem cell renewal during somatic embryogenesis in Arabidopsis." Plant J **59**(3): 448-460.

Suzuki, M. and D. R. McCarty (2008). "Functional symmetry of the B3 network controlling seed development." Current Opinion in Plant Biology **11**(5): 548-553.

Suzuki, M., H. H. Y. Wang and D. R. McCarty (2007). "Repression of the LEAFY COTYLEDON 1/B3 regulatory network in plant embryo development by VP1/ABSCISIC ACID INSENSITIVE 3-LIKE B3 genes." Plant Physiol **143**(2): 902-911.

Szczygiel-Sommer, A. and M. D. Gaj (2019). "The miR396-GRF regulatory module controls the embryogenic response in Arabidopsis via an auxin-related pathway." International Journal of Molecular Sciences **20**(20): article no. 5221.

Szyrajew, K., D. Bielewicz, J. Dolata, A. M. Wojcik, K. Nowak, A. Szczygiel-Sommer, Z. Szweykowska-Kulinska, A. Jarmolowski and M. D. Gaj (2017). "MicroRNAs are intensively regulated during induction of somatic embryogenesis in Arabidopsis." Frontiers in Plant Science **8**: article no. 18.

Tanaka, M., A. Kikuchi and H. Kamada (2008). "The Arabidopsis histone deacetylases HDA6 and HDA19 contribute to the repression of embryonic properties after germination." Plant Physiol **146**(1): 149-161.

Tang, L. P., C. Zhou, S. S. Wang, J. Yuan, X. S. Zhang and Y. H. Su (2017). "FUSCA3 interacting with LEAFY COTYLEDON2 controls lateral root formation through regulating *YUCCA4* gene expression in Arabidopsis thaliana." New Phytologist **213**(4): 1740-1754.

Tao, Z., H. M. Hu, X. Luo, B. Jia, J. M. Du and Y. H. He (2019). "Embryonic resetting of the parental vernalized state by two B3 domain transcription factors in Arabidopsis." Nature Plants **5**(4): 424-435.

Tao, Z., L. S. Shen, X. F. Gu, Y. Z. Wang, H. Yu and Y. H. He (2017). "Embryonic epigenetic reprogramming by a pioneer transcription factor in plants." Nature **551**(7678): 124-128.

Thakare, D., W. Tang, K. Hill and S. E. Perry (2008). "The MADS-domain transcriptional regulator AGAMOUS-LIKE15 promotes somatic embryo development in Arabidopsis and soybean." Plant Physiol **146**(4): 1663-1672.

Thomas, T. L. (1993). "Gene expression during plant embryogenesis and germination: An overview." Plant Cell **5**(10): 1401-1410.

Tian, R., F. Wang, Q. Zheng, V. Niza, A. B. Downie and S.E.Perry (2020). "Direct and indirect targets of the Arabidopsis seed transcription factor ABSCISIC ACID INSENSITIVE3." Plant J : in press.

To, A., C. Valon, G. Savino, J. Guilleminot, M. Devic, J. Giraudat and F. Parcy (2006). "A network of local and redundant gene regulation governs *Arabidopsis* seed maturation." Plant Cell **18**: 1642-1651.

Tsuchiya, Y., E. Nambara, S. Naito and P. McCourt (2004). "The FUS3 transcription factor functions through the epidermal regulator TTG1 during embryogenesis in Arabidopsis." Plant J **37**(1): 73-81.

Tsuwamoto, R., S. Yokoi and Y. Takahata (2010). "Arabidopsis EMBRYOMAKER encoding an AP2 domain transcription factor plays a key role in developmental change from vegetative to embryonic phase." Plant Molecular Biology **73**(4-5): 481-492.

Tuteja, N. J. (2007). "Abscisic acid and abiotic stress signaling." Plant Signaling & Behavoir **2**(3): 135-138.

Ufaz, S., V. Shukla, Y. Soloveichik, Y. Golan, F. Breuer, Z. Koncz, G. Galili, C. Koncz and A. Zilberstein (2011). "Transcriptional control of aspartate kinase expression during darkness and sugar depletion in Arabidopsis: involvement of bZIP transcription factors." Planta **233**(5): 1025-1040.

Ullah, A., H. Manghwar, M. Shaban, A. H. Khan, A. Akbar, U. Ali, E. Ali and S. Fahad (2018). "Phytohormones enhanced drought tolerance in plants: A coping strategy." Environmental Science Pollution Research **25**(33): 33103-33118.

Veerappan, V., J. Wang, M. Y. Kang, J. Lee, Y. H. Tang, A. K. Jha, H. Z. Shi, R. Palanivelu and R. D. Allen (2012). "A novel HSI2 mutation in Arabidopsis affects the PHD-like domain and leads to derepression of seed-specific gene expression." Planta **236**(1): 1-17.

Vernoud, V., M. Hajduch, A. Khaled, N. Depège and M. Rogowsky (2005). "Maize embryogenesis." Maydica **50**(3/4): 469-483.

Vinocur, B. and A. Altman (2005). "Recent advances in engineering plant tolerance to abiotic stress: achievements and limitations." Current opinion in biotechnology **16**(2): 123-132.

Vishwakarma, K., N. Upadhyay, N. Kumar, G. Yadav, J. Singh, R. K. Mishra, V. Kumar, R. Verma, R. Upadhyay and M. Pandey (2017). "Abscisic acid signaling and abiotic stress tolerance in plants: a review on current knowledge and future prospects." Frontiers in Plant Science **8**: 161.

Vurukonda, S. S. K. P., S. Vardharajula, M. Shrivastava and A. SkZ (2016). "Enhancement of drought stress tolerance in crops by plant growth promoting rhizobacteria." Microbiological research **184**: 13-24.

Waese, J., J. Fan, A. Pasha, H. Yu, G. Fucile, R. Shi, M. Cumming, L. A. Kelley, M. J. Sternberg, V. Krishnakumar, E. Ferlanti, J. Miller, C. Town, W. Stuerzlinger and N. J. Provart (2017). "ePlant: Visualizing and Exploring Multiple Levels of Data for Hypothesis Generation in Plant Biology." Plant Cell **29**(8): 1806-1821.

Waki, T., T. Hiki, R. Watanabe, T. Hashimoto and K. Nakajima (2011). "The Arabidopsis RWP-RK protein RKD4 triggers gene expression and pattern formation in early embryogenesis." Current Biology **21**(15): 1277-1281.

Wang, F. F. and S. E. Perry (2013). "Identification of direct targets of FUSCA3, a key regulator of Arabidopsis seed development." Plant Physiol **161**(3): 1251-1264.

Wang, H., L. V. Caruso, A. B. Downie and S. E. Perry (2004). "The embryo MADS domain protein AGAMOUS-Like 15 directly regulates expression of a gene encoding an enzyme involved in gibberellin metabolism." Plant Cell **16**(5): 1206-1219.

Wang, H., W. Tang, C. Zhu and S. E. Perry (2002). "A chromatin immunoprecipitation (ChIP) approach to isolate genes regulated by AGL15, a MADS-domain protein that preferentially accumulates in embryos." Plant J **32**: 831-843.

Wang, K., H. M. Chen, Y. J. Miao and M. Bayer (2020). "Square one: zygote polarity and early embryogenesis in flowering plants." Current Opinion in Plant Biology **53**: 128-133.

Wang, X. C., Q. W. Niu, C. Teng, C. Li, J. Y. Mu, N. H. Chua and J. R. Zuo (2009). "Overexpression of *PGA37/MYB118* and *MYB115* promotes vegetative-to-embryonic transition in Arabidopsis." Cell Research **19**(2): 224-235.

Wendrich, J. R. and D. Weijers (2013). "The Arabidopsis embryo as a miniature morphogenesis model." New Phytologist **199**(1): 14-25.

West, M. and J. J. Harada (1993). "Embryogenesis in higher plants: an overview." Plant Cell **5**(10): 1361-1369.

Willmann, M. R., A. J. Mehalick, R. L. Packer and P. D. Jenik (2011). "MicroRNAs regulate the timing of embryo maturation in Arabidopsis." Plant Physiol **155**(4): 1871-1884.

Winter, D., B. Vinegar, H. Nahal, R. Ammar, G. V. Wilson and N. J. Provart (2007). "An "electronic fluorescent pictograph" browser for exploring and analyzing large-scale biological data sets." Plos One **2**(8): article no. e718.

Wise, M. J. (2003). "LEAping to conclusions: a computational reanalysis of late embryogenesis abundant proteins and their possible roles." BMC bioinformatics **4**(1): 52.

Wojcik, A. M., B. Wojcikowska and M. D. Gaj (2020). "Current perspectives on the auxin-mediated genetic network that controls the induction of somatic embryogenesis in plants." International Journal of Molecular Sciences **21**(4): article no. 1333.

Wojcikowska, B. and M. D. Gaj (2015). "LEAFY COTYLEDON2-mediated control of the endogenous hormone content: implications for the induction of somatic embryogenesis in Arabidopsis." Plant Cell Tissue and Organ Culture **121**(1): 255-258.

Wojcikowska, B., K. Jaskola, P. Gasiorek, M. Meus, K. Nowak and M. D. Gaj (2013). "LEAFY COTYLEDON2 (LEC2) promotes embryogenic induction in somatic tissues of Arabidopsis, via YUCCA-mediated auxin biosynthesis." Planta **238**(3): 425-440.

Xi, W. Y., C. Liu, X. L. Hou and H. Yu (2010). "MOTHER OF FT AND TFL1 Regulates Seed Germination through a Negative Feedback Loop Modulating ABA Signaling in Arabidopsis." Plant Cell **22**(6): 1733-1748.

Xiang, C. B., P. Han and D. J. Oliver (1999). "In solium selection for arabidopsis transformants resistant to kanamycin." Plant Molecular Biology Reporter **17**(1): 59-65.

Yamamoto, A., Y. Kagaya, R. Toyoshima, M. Kagaya, S. Takeda and T. Hattori (2009). "Arabidopsis NF-YB subunits LEC1 and LEC1-LIKE activate transcription by interacting with seed-specific ABRE-binding factors." Plant J **58**(5): 843-856.

Yamamoto, A., Y. Kagaya, H. Usui, T. Hobo, S. Takeda and T. Hattori (2010). "Diverse roles and mechanisms of gene regulation by the Arabidopsis seed maturation master regulator FUS3 revealed by microarray analysis." Plant and Cell Physiology **51**(12): 2031-2046.

Yan, W., D. Chen and K. Kaufmann (2016). "Efficient multiplex mutagenesis by RNA-guided Cas9 and its use in the characterization of regulatory elements in the AGAMOUS gene." Plant methods **12**(1): 23.

Yang, Y., J. Fear, J. Hu, I. Haecker, L. Zhou, R. Renne, D. Bloom and L. M. McIntyre (2014). "Leveraging biological replicates to improve analysis in ChIP-seq experiments." Computational structural biotechnology journal **9**(13): e201401002.

Yang, Z. R., C. F. Li, Y. Wang, C. J. Zhang, Z. X. Wu, X. Y. Zhang, C. L. Liu and F. G. Li (2014). "GhAGL15s, preferentially expressed during somatic embryogenesis, promote embryogenic callus formation in cotton (Gossypium hirsutum L.)." Molecular Genetics and Genomics **289**(5): 873-883.

Yoshida, S., P. B. de Reuille, B. Lane, G. W. Bassel, P. Prusinkiewicz, R. S. Smith and D. Weijers (2014). "Genetic control of plant development by overriding a geometric division rule." Developmental Cell **29**(1): 75-87.

Zhang, H., B. Bishop, W. Ringenberg, W. M. Muir and J. Ogas (2012). "The CHD3 Remodeler PICKLE Associates with Genes Enriched for Trimethylation of Histone H3 Lysine 27." Plant Physiol **159**(1): 418-432.

Zhang, H., S. D. Rider, J. T. Henderson, M. Fountain, K. Chuang, V. Kandachar, A. Simons, H. J. Edenberg, J. Romero-Severson and W. M. Muir (2008). "The CHD3 remodeler PICKLE promotes trimethylation of histone H3 lysine 27." Journal of Biological Chemistry **283**(33): 22637-22648.

Zhang, Y., G. Cao, L.-J. Qu and H. J. P. C. R. Gu (2009). "Involvement of an R2R3-MYB transcription factor gene AtMYB118 in embryogenesis in Arabidopsis." **28**(3): 337-346.

Zhao, P., K. Begcy, T. Dresselhaus and M.-X. Sun (2017). "Does early embryogenesis in eudicots and monocots involve the same mechanism and molecular players?" Plant Physiol **173**(1): 130-142.

Zheng, Q. L. and S. E. Perry (2014). "Alterations in the transcriptome of soybean in response to enhanced somatic embryogenesis promoted by orthologs of AGAMOUS-Like15 and AGAMOUS-Like18." Plant Physiol **164**(3): 1365-1377.

Zheng, Y., N. Ren, H. Wang, A. J. Stromberg and S. E. Perry (2009). "Global identification of targets of the Arabidopsis MADS domain protein AGAMOUS-Like 15." Plant Cell **21**: 2563-2577.

Zhou, Y., B. Tan, M. Luo, Y. Li, C. Liu, C. Chen, C. W. Yu, S. G. Yang, S. Dong, J. X. Ruan, L. B. Yuan, Z. Zhang, L. M. Zhao, C. L. Li, H. H. Chen, Y. H. Cui, K. Q. Wu and S. Z. Huang (2013). "HISTONE DEACETYLASE19 interacts with HSL1 and participates in the repression of seed maturation genes in Arabidopsis seedlings." Plant Cell **25**(1): 134-148.

Zhu, S. P., J. Wang, J. L. Ye, A. D. Zhu, W. W. Guo and X. X. Deng (2014). "Isolation and characterization of LEAFY COTYLEDON 1-LIKE gene related to embryogenic competence in Citrus sinensis." Plant Cell Tissue and Organ Culture **119**(1): 1-13.

Zuo, J. R., Q. W. Niu, G. Frugis and N. H. Chua (2002). "The *WUSCHEL* gene promotes vegetative-to-embryonic transition in Arabidopsis."Plant J **30**(3): 349-359.

**Vita**

**Ran Tian**

**Place of birth:**

Guizhou, China

**Educational institutions:**

B.S. Hebei Normal University, Hebei, China, 2009-2013

M.S. University of Louisville, Kentucky, U.S, 2013-2015

Ph.D. University of Kentucky, Kentucky, U.S 2015-present

**Professional positions held:**

2010, 2012 Intern over Summer Vacation, the Institute of Cultural Relics and Archaeology, China

2012-2013 Research Technician, Hebei Normal University, China

2014-2015 Research Assistant, University of Louisville, U.S

2017-2018 Teaching Assistant, University of Kentucky, U.S

2015-present Research Assistant, University of Kentucky, U.S

**Scholastic and professional honors:**

2010 The Excellent Prize in Summer Vacation Internship.

2011 *The Genetic Screening of Acidophilic Hot Alicyclobacillus bacillus Alanine Racemase,* 3rd Prize in the Scientific and Technological Innovation Competition.

2011 The 3rd Scholarship awarded by Hebei Normal University.

2012 *Small Scale Plasmid DNA Preparation*, Published on Guizhou University Journal, 2012 Supplement.

2017 Travel funds from the graduate school and the department for attending the International Conference on Arabidopsis Research at the St. Louis Arch, St. Louis to present a poster.

2019 TeKrony Graduate Student Travel Fellowship in Seed Biology for attending the The Annual Plant Biology Meeting at San Jose, California to present a poster.

**Professional publications:**

**Tian, R.**, F. Wang, Q. Zheng, V. Niza, A. B. Downie and S.E.Perry (2020). "Direct and indirect targets of the Arabidopsis seed transcription factor ABSCISIC ACID INSENSITIVE3." Plant J : in press.

Signature Ran Tian