Transcriptomic Signature of the SHATTERPROOF2 Expression Domain Reveals the Meristematic Nature of Arabidopsis Gynoecial Medial Domain

Gonzalo H. Villarino, Qiwen Hu, Silvia Manrique, Miguel Flores-Vergara, Bhupinder Sehra, Linda Robles, Javier Brumos, Anna N. Stepanova, Lucia Colombo, Eva Sundberg, Steffen Heber, and Robert G. Franks*

Department of Plant and Microbial Biology (G.H.V., M.F.-V., B.S., L.R., J.B., A.N.S., R.G.F.) and Department of Computer Science and Bioinformatics Research Center (Q.H., S.H.), North Carolina State University, Raleigh, North Carolina 27606; Università degli Studi di Milano Dip. di BioScienze, Sezione di Botanica Generale, Milan, Italy 20133 (S.M., L.C.); and Department of Plant Biology, Swedish University of Agricultural Sciences, Uppsala, Sweden 750 07 (E.S.)

ORCID IDs: 0000-0001-6241-4143 (S.M.); 0000-0002-7815-5905 (L.R.); 0000-0001-8415-1399 (L.C.); 0000-0002-3978-1030 (R.G.F.).

Plant meristems, like animal stem cell niches, maintain a pool of multipotent, undifferentiated cells that divide and differentiate to give rise to organs. In Arabidopsis (Arabidopsis thaliana), the carpel margin meristem is a vital meristematic structure that generates ovules from the medial domain of the gynoecium, the female floral reproductive structure. The molecular mechanisms that specify this meristematic region and regulate its organogenetic potential are poorly understood. Here, we present a novel approach to analyze the transcriptional signature of the medial domain of the Arabidopsis gynoecium, highlighting the developmental stages that immediately precede ovule initiation, the earliest stages of seed development. Using a floral synchronization system and a SHATTERPROOF2 (SHP2) domain-specific reporter, paired with FACS and RNA sequencing, we assayed the transcriptome of the gynoecial medial domain with temporal and spatial precision. This analysis reveals a set of genes that are differentially expressed within the SHP2 expression domain, including genes that have been shown previously to function during the development of medial domain-derived structures, including the ovules, thus validating our approach. Global analyses of the transcriptomic data set indicate a similarity of the pSHP2-expressing cell population to previously characterized meristematic domains, further supporting the meristematic nature of this gynoecial tissue. Our method identifies additional genes including novel isoforms, cis-natural antisense transcripts, and a previously unrecognized member of the REPRODUCTIVE MERISTEM family of transcriptional regulators that are potential novel regulators of medial domain development. This data set provides genome-wide transcriptional insight into the development of the carpel margin meristem in Arabidopsis.

The seed pod of flowering plants develops from the gynoecium, the female reproductive structure of the flower (Seymour et al., 2013). The gynoecium generates the ovules (the precursors of the seeds) and develops into the edible fruit in many fruiting species. As an estimated two-thirds of the calories of humankind’s diet are derived from gynoecia and seeds, the gynoecium is a globally vital structure (Oram and Brock, 1972; Singh and Bhalla, 2007).

In the flowering plant Arabidopsis (Arabidopsis thaliana), the gynoecium is a morphologically complex, multiorgan structure with a diversity of tissues and cell types (Sessions and Zambrjksi, 1995; Bowman et al., 1999; Seymour et al., 2013). The mature gynoecium displays morphological and functional differentiation along apical–basal, medio–lateral, and adaxial–abaxial (inner–outer) axes. Stigmatic and stylar tissue form at the apex of the gynoecium, where the pollen grains are received and germinate. The stigma and style also constitute the apical-most portion of the transmitting tract, a structure that allows the pollen tube cell and sperm cells to reach the internally located female gametophytes (Sessions and Zambrjksi, 1995; Sessions, 1999; Crawford and Yanofsky, 2008, Fig. 1, A–C). Located basal to the stigmatic and stylar tissue is the ovary portion of the gynoecium.

Ovules form within the ovary from a meristematic structure termed the medial ridge or carpel margin
meristem (CMM), located in medial portions of the gynoecium (Bowman et al., 1999; Alvarez and Smyth, 2002; Reyes-Olalde et al., 2013; Fig. 1, A–C). Plant meristems are analogous to animal stem cell niches, as they maintain a set of undifferentiated cells that can divide and differentiate into numerous tissues and cell types (Aichinger et al., 2012). Early during floral development, patterning events divide the gynoecial primordium into a medial domain that contains the CMM and lateral domains that will form the walls of the gynoecium (Bowman et al., 1999). These domains express different sets of transcriptional regulators from early developmental time points.

Many genes that play a role in the development of the CMM and in the generation of ovules from this structure have been analyzed previously (Reyes-Olalde et al., 2013). However, due to the complexity of the developing gynoecium and the heterogeneity of the gynoecial tissues, the ability to analyze the transcriptomic signature of the developing CMM or even other specific developing gynoecial structural domains has been limited. Wynn et al. (2011) previously evaluated the transcriptional properties of the gynoecial medial domain using hand-dissected gynoecial samples from the *seu* antegumenta (*seu* ant) double mutants that display a loss of many medial domain-derived structures, including ovules. They identified 210 genes displaying reduced expression in *seu* ant gynoecia from floral stages 8 to 10 (stages according to Smyth et al. [1990]). Many of these genes were shown via in situ hybridization to be preferentially expressed in the developing medial domain of the wild-type gynoecium, and several of these genes have been shown to function during the development of ovules from the medial domain (Reyes-Olalde et al., 2013). However, it is difficult with this approach to obtain samples from gynoecia younger than stage 8 and, thus, to assay the earliest gynoecial patterning events.

An alternative approach to investigate the transcriptional properties of specific cellular populations utilizes FACS of protoplasted cells to isolate specific cell populations based on patterns of gene expression. This approach has been applied successfully to the Arabidopsis shoot apical meristem (SAM; Yadav et al., 2009, 2014) and roots (Birnbaum et al., 2003, 2005;
Carter et al., 2013; Lan et al., 2013) as well as to developing cell lineages within the Arabidopsis leaf epidermis (Adrian et al., 2015).

Here, we developed a novel FACS-based system for the transcriptomic analysis of a specific cellular population from the developing gynoecium, specifically the population of cells expressing the transcriptional regulator SHATTERPROOF2 (SHP2). SHP2 encodes a MADS domain transcription factor that is expressed early within the developing CMM and, thus, functions as a marker for the meristematic population of cells that generate the transmitting tract and ovules (Ma et al., 1991; Savidge et al., 1995; Colombo et al., 2010; Larsson et al., 2014). In order to focus our analysis on early stages of gynoecium development during which key patterning events occur, we generated a SHP2 domain-specific reporter in a genetic background that allowed the synchronization of floral development. This, coupled with FACS-based protoplast sorting procedures and RNA sequencing (RNA-seq), provided a unique temporal and spatial precision to assay the transcriptional signature of the gynoecial SHP2 expression domain.

Our system provides the ability to isolate large numbers of cells from a temporally and spatially restricted gynoecial domain. We apply this method to investigate the transcriptomic signature of the medial domain of the gynoecium at the developmental stages when key patterning events and ovule initiation occur. Our analysis reveals many genes that are expressed preferentially within the developing medial portions of the gynoecium, including members of the REPRODUCTIVE MERISTEM (REM) family of transcriptional regulators (Swaminathan et al., 2008; Romanel et al., 2009). We also take advantage of strand-specific RNA-seq technology to find protein coding genes and noncoding RNAs as well as to examine isoforms and naturally occurring antisense transcripts that are expressed preferentially in the medial domain. This work complements and extends previous analyses of medial domain development and generates a list of potential novel regulators of medial domain development that are strong candidates for future functional analyses. Furthermore, global analyses of the transcriptomic data set indicate a similarity of the pSHP2-expressing cell population to previously characterized meristematic domains, further supporting the meristematic nature of this gynoecial tissue.

RESULTS AND DISCUSSION

FACS-Based Protoplast Sorting Allows the Collection of the SHP2-Expressing Cell Population from a Temporally Restricted Inflorescence Sample

The transcriptional regulator SHP2 is preferentially expressed in the medial domain of the gynoecium and in a subset of the medial domain-derived tissues (Savidge et al., 1995; Colombo et al., 2010; Larsson et al., 2014; Fig. 1, D and E). SHP2 plays an important role in the development of the medial domain and in the specification of ovule identity (Liljegren et al., 2000; Favaro et al., 2003; Pinyopich et al., 2003; Colombo et al., 2010; Galbiati et al., 2013). To better characterize the molecular mechanisms of the medial domain and ovule development, we sought to identify transcripts that are differentially expressed within the medial domain of the Arabidopsis gynoecium relative to the rest of the inflorescence. To enable this, we generated a transgenic line containing a two-component reporter system in which a pUAS-3xYPET reporter was driven by a pSHP2-GAL4 driver construct (see “Materials and Methods”). Throughout this article, we refer to this two-component reporter as pSHP2-YFP (YFP for yellow fluorescent protein).

To better understand the early specification of medial and lateral gynoecial domains and in the earliest stages of ovule primordium initiation, we focused our transcriptomic analysis on floral stages 6 to 8, when these key developmental events occur (Bowman et al., 1999). In order to increase our ability to collect a large number of pSHP2-YFP-expressing cells from this specific bracket of developmental stages, we crossed the pSHP2-YFP reporter into an ap1 cal-based floral synchronization system that allows the collection of large numbers of semisynchronized flowers at roughly the same developmental stage (Wellmer et al., 2006; O’Maoléidigh and Wellmer, 2014). The expression of the pSHP2-YFP reporter in the floral synchronization system was largely similar to that observed in wild-type inflorescences (Ma et al., 1991; Colombo et al., 2010; Larsson et al., 2014) and was confined chiefly to the medial domain and medial domain-derived tissues (Fig. 1, D and E). Some expression was observed in nonmedial gynoecial positions. The most apparent of this was within the apex of the developing gynoecium, where the pSHP2-YFP reporter is detected in both medial and lateral positions. Additionally, expression could be observed in a small number of cells within the stamens (Fig. 1D) and occasionally on the edges of sepals that appeared to have undergone a homeotic transformation toward a carpelloid fate (data not shown). Thus, the vast majority of the pSHP-YFP reporter expression reflected the endogenous pSHP2 expression domain (in the medial and apical portions of the gynoecium). We cannot exclude the possibility that a small minority of the expression domain may reflect ectopic expression of the reporter due to the genetic background or transgene insertion site or limitations of the regulatory sequences used in the pSHP-YFP reporter construct.

Microscopic examination of our semisynchronized inflorescence samples indicated that flowers ranged between floral stage 1 and early stage 8, with a strong enrichment for floral stages 6 through early 8 (Fig. 1, G and H). Flowers that had developed beyond late stage 8 were not detected in our samples. Thus, our biological sample is strongly enriched for transcripts that are expressed during early patterning of the gynoecium and the earliest stages of ovule development (initiation).
and does not include later floral developmental stages, where SHP2 is expressed in stigma, style, and valve margin tissues. Additionally, as the initial expression of the pSHP-YFP reporter is detected at late stage 5 or early stage 6 (Larsson et al., 2014), we expect that the population of YFP-expressing protoplasts derived from this material will be highly enriched with cells from the stage 6 to 8 medial domain.

FACS sorting of protoplasts derived from these inflorescences yielded three populations of sorted cells (collected in biological quadruplicate): YFP-positive, YFP-negative, and all-sorted cells (Supplemental Fig. S1). The all-sorted sample included all protoplasts recovered (regardless of YFP expression) after sorting gates were applied to remove debris and broken cells (see “Materials and Methods”). We additionally collected (also in biological quadruplicate) nonsorted samples from entire nonprotoplasted inflorescences to measure the abundance of transcripts in the biological starting material before protoplast generation and FACS sorting. In order to evaluate the purity of the YFP-positive protoplasts during a preliminary FACS run, YFP-positive cells were resorted. Ninety-six percent of the YFP-positive cells were found to sort into the YFP-positive gate, indicating a high degree of enrichment and purity in the YFP-positive sample (Supplemental Fig. S1). Confocal microscopy also revealed an enriched population of intact YFP-positive protoplasts after FACS (Fig. 1I).

We used quantitative real-time (qRT)-PCR to estimate the degree of enrichment of the endogenous SHP2 and NGATHA1 (NGA1) transcripts in RNA samples derived from the YFP-positive and YFP-negative samples. NGA1 is expressed in the adaxial portions of the gynoecium starting at stage 7 in a domain that partially overlaps with the SHP2 expression domain (Alvarez et al., 2009; Trigueros et al., 2009) and thus provides an additional benchmark to estimate the enrichment of medial domain-expressed transcripts. The normalized expression of the SHP2 transcript was approximately 30-fold higher in the YFP-positive samples relative to the YFP-negative samples ($P = 0.001$), while the NGA1 transcript was approximately 4-fold higher in the YFP-positive samples ($P = 0.05$). The difference in the levels of TUBULIN6 (TUB6) was not found to be statistically significant ($P = 0.4$) between the YFP-positive and YFP-negative samples (Supplemental Fig. S2).

Transcriptomic Analysis of the Gynoecial SHP2 Expression Domain and Identification of Candidate Regulators of Gynoecial Medial Domain Development

To investigate the transcriptomic profile of the gynoecial SHP2 expression domain, we performed high-throughput RNA-seq from the collected protoplasts and nonprotoplasted inflorescence samples. We expect that the identification of differentially expressed genes (DEGs) between the YFP-positive and YFP-negative samples (referred to as YFP+/−) will provide insight into the set of transcripts differentially expressed in the gynoecial medial domain relative to the rest of the inflorescence. Additionally, DEGs identified in the all-sorted and nonsorted comparison (referred to as all-sorted/ nonsorted) are expected to reveal transcripts that are differentially represented as a result of the protoplasting/FACS-sorting protocol.

Two lanes of the HiSeq2500 Illumina sequencing platform yielded 320 million raw reads, with an average of 20 million reads per library. Nearly 11 million reads were filtered out after removing barcode adapters and low-quality sequences. The remaining 306 million reads were aligned against The Arabidopsis Information Resource (TAIR) 10 reference genome (Lamesch et al., 2012), with more than 90% of them successfully mapping to the genome sequence. Among the mapped reads, 244 million reads mapped uniquely to only one location and were used for subsequent analyses. A detailed breakdown is shown in Supplemental Table S1.

We used three different programs to determine expressed and differentially expressed protein-coding genes in our data set: Cufflinks (Trapnell et al., 2012), edgeR (Robinson et al., 2010), and DESeq2 (Love et al., 2014; see “Materials and Methods”); non-protein-coding gene models were considered separately and are presented below. These open-source programs are among the most cited RNA-seq analysis tools used for the identification of DEGs. Various studies suggest that, among the different RNA-seq analysis tools, no single method outperforms the other ones (Rapaport et al., 2013; Soneson and Delorenzi, 2013). In our study, we compared the results of edgeR, DESeq2, and Cufflinks using our data set (Fig. 2). Our choice of these three tools reflects significant differences in the underlying algorithms. Cufflinks uses isoform expression estimates, while edgeR and DESeq2 compute gene level-based expression estimates. Furthermore, edgeR and DESeq2 use different normalization approaches and dispersion estimates. As these three programs use different algorithms, we expected each program to return a somewhat different set of DEGs. Here, the term DEG is used to indicate a gene whose steady-state transcript level differs significantly at a false discovery rate (FDR) < 0.001 and shows a fold change of 4 or more between the two compared RNA samples.

We reasoned that any gene identified as a DEG by all three programs was highly likely to be differentially expressed in our data set. Thus, to identify potential regulators of gynoecial medial domain development, a stringent criterion was used to select a subset of the YFP+/− DEGs for downstream analysis. For a gene to be selected from the YFP+/− comparison, we required that the transcript be identified as differentially expressed by all three independent software packages. This led to the identification of 411 triply identified DEGs (e.g. the overlap sections in the Venn diagram in Fig. 2B).

Additionally, we sought to confirm that these 411 genes were not differentially expressed as a result of the protoplasting and sorting procedure. We reasoned that DEGs identified as differentially expressed between the all-sorted and nonsorted samples would characterize
leaving 363 cleaned protein-coding DEGs from the
protoplast generation or FACS-sorting procedures, these 48 transcripts from our analysis to eliminate any
YFP+/− DEGs were found in common between the
(Fig. 2C), indicating a high degree of speci-
entially expressed as a result of the protoplasting
expression profiles. B, Venn diagram showing DEGs between YFP+/− samples identified in the three programs. C, Intersection of the DEGs (48) from both
data sets (A and B). A total of 363 DEGs, after removing DEGs induced by the protoplasting/FACS-sorting stress, were used for downstream analysis.

For the 95 DEGs that were enriched in the YFP-positive sample (at a fold change > 4), we expected many to be preferentially expressed in the medial portions of the gynoecium at floral stages 6 to 8. To test this, we examined the literature to determine the expression patterns of members of this gene set. From the top 15 of the 95 YFP-positive enriched DEGs (ranked by fold change), five have been reported previously to be preferentially expressed in the gynoecial medial domain via in situ or reporter gene analysis (i.e. HECATE1 [HEC1], HEC2, SHP1, SHP2, and STYLISH1 [STY1]; Ma et al., 1991; Savidge et al., 1995; Kuusk et al., 2002; Gremski et al., 2007; Colombo et al., 2010) and three others were described previously as enriched in medial domain-derived tissues in published transcriptomic data sets (i.e. AT1G66950, AT5G14180, and AT1G03720; Skinner and Gasser, 2009; Wuest et al., 2010; Table I).

An additional gene from this list, CRABS CLAW (CRC), has been shown via in situ hybridization to be expressed in portions of the medial gynoecial domain as well as nonmedial portions of the gynoecium (Bowman and Smyth, 1999; Åzakanandam et al., 2008). The expression pattern of the remaining six genes from this gene list have not yet been assayed in the gynoecium. Thus, as predicted, the set of 95 genes enriched in the YFP-positive sample is enriched for genes that are preferentially expressed in the gynoecial medial domain. Published functional analyses of HEC1, HEC2, SHP1, SHP2, and STY1 indicate that these genes function during the development of the medial domain or medial domain-derived tissues (Kuusk et al., 2002; Favaro et al., 2003; Pinyopich et al., 2003; Gremski et al., 2007; Colombo et al., 2010). Many other genes in the set of 95 DEGs enriched in the YFP-positive sample have been shown previously to play a role in medial domain development (e.g. NGA family members [Alvarez et al., 2009; Trigueros et al., 2009], SPT [Heisler et al., 2001], and CUC2 [Kamiuchi et al., 2014]). Other genes within this list are interesting candidates for future functional studies. These include members of the REM family of transcriptional regulators (Swaminathan et al., 2008; Romanel et al., 2009), several auxin synthesis- or signaling-related genes, such as LIKE AUXIN RESISTANT1 (LAX1 [AT5G01240]; Bennett et al., 1996) and YUCCA4 (YUC4 [AT5G11320]; Cheng et al., 2006), as well as transcription factors regulating other developmental processes, such as MATERNAL EFFECT EMBRYO ARREST3 (MEE3 [AT2G21650]; Pagnussat et al., 2005) and GLABROUS3 (AT5G41315; Payne et al., 2000).

It is important to note that the 48 DEGs that were identified in both the YFP+/− and all-sorted/nonsorted comparisons (Fig. 2C) should not be discounted as potential medial domain regulators. These genes may be both preferentially expressed in the YFP-positive cell population and induced in response to the protoplasting procedure (Supplemental Table S2). Indeed, some of these genes, including the transcription factors HEC3 and BRASSINOSTEROID-ENHANCED EXPRESSION1, have been reported to be preferentially expressed in medial domain-derived tissues and to function in...
Domain-Specific Transcriptomic Analysis

Table I. Top 15 DEGs enriched in the YFP-positive sample as ranked by fold change

Arabidopsis gene identifiers are shown in the first column, gene names (TAIR 10 annotation) are shown in the second column, and the third column shows the reference for each available reporter line and/or in situ hybridization (NA, not applicable). Average RPKM values are indicated for each sample (YFP_NEG = YFP negative and YFP_POS = YFP positive).

<table>
<thead>
<tr>
<th>Gene Identifier</th>
<th>Gene Name</th>
<th>Reference for Expression</th>
<th>YFP_NEG Average Expression</th>
<th>YFP_POS Average Expression</th>
<th>Fold Expression Difference</th>
<th>P</th>
<th>FDR</th>
</tr>
</thead>
<tbody>
<tr>
<td>AT3G50330</td>
<td>HEC2</td>
<td>Gremski et al. (2007)</td>
<td>1.267</td>
<td>80.148</td>
<td>67.0</td>
<td>1.26E-31</td>
<td>5.36E-18</td>
</tr>
<tr>
<td>AT5G67060</td>
<td>HEC1</td>
<td>Gremski et al. (2007)</td>
<td>0.708</td>
<td>40.298</td>
<td>59.4</td>
<td>1.44E-105</td>
<td>2.44E-101</td>
</tr>
<tr>
<td>AT1G66950</td>
<td>ATPDR11</td>
<td>Wuest et al. (2010)</td>
<td>0.136</td>
<td>5.931</td>
<td>45.2</td>
<td>9.47E-76</td>
<td>2.02E-72</td>
</tr>
<tr>
<td>AT3G58780</td>
<td>SHP1</td>
<td>Ma et al. (1991)</td>
<td>1.678</td>
<td>54.233</td>
<td>33.9</td>
<td>1.06E-32</td>
<td>7.54E-50</td>
</tr>
<tr>
<td>AT5G17040</td>
<td>Unknown</td>
<td>NA</td>
<td>0.272</td>
<td>7.453</td>
<td>28.0</td>
<td>1.06E-29</td>
<td>1.24E-27</td>
</tr>
<tr>
<td>AT1G06920</td>
<td>OFFP</td>
<td>NA</td>
<td>0.453</td>
<td>11.267</td>
<td>25.9</td>
<td>6.18E-34</td>
<td>1.04E-31</td>
</tr>
<tr>
<td>AT3G51060</td>
<td>STY1</td>
<td>Kuusk et al. (2002)</td>
<td>3.141</td>
<td>73.227</td>
<td>24.5</td>
<td>2.22E-98</td>
<td>1.89E-94</td>
</tr>
<tr>
<td>AT2G42830</td>
<td>SHP2</td>
<td>Ma et al. (1991)</td>
<td>1.642</td>
<td>31.392</td>
<td>20.1</td>
<td>1.58E-74</td>
<td>2.99E-71</td>
</tr>
<tr>
<td>AT1G03720</td>
<td>Unknown</td>
<td>Skinner and Gasser (2009)</td>
<td>1.898</td>
<td>31.968</td>
<td>17.7</td>
<td>6.66E-45</td>
<td>4.54E-49</td>
</tr>
<tr>
<td>AT2G33850</td>
<td>Unknown</td>
<td>NA</td>
<td>0.676</td>
<td>10.904</td>
<td>16.7</td>
<td>1.85E-19</td>
<td>7.15E-18</td>
</tr>
<tr>
<td>AT1G69180</td>
<td>CRC</td>
<td>Bowman et al. (1999)</td>
<td>19.435</td>
<td>296.675</td>
<td>16.0</td>
<td>2.60E-78</td>
<td>8.84E-75</td>
</tr>
<tr>
<td>AT2G22460</td>
<td>Unknown</td>
<td>NA</td>
<td>0.727</td>
<td>10.899</td>
<td>15.6</td>
<td>2.85E-17</td>
<td>8.52E-16</td>
</tr>
<tr>
<td>AT2G21650</td>
<td>MEE3</td>
<td>Pagnussat et al. (2005); Baxter et al. (2007)</td>
<td>1.333</td>
<td>17.643</td>
<td>13.9</td>
<td>1.65E-16</td>
<td>4.37E-15</td>
</tr>
<tr>
<td>AT2G24540</td>
<td>AFR</td>
<td>NA</td>
<td>0.84</td>
<td>11.029</td>
<td>13.8</td>
<td>2.27E-24</td>
<td>1.61E-22</td>
</tr>
</tbody>
</table>

REM Family Members Are Differentially Expressed in the SHP2 Expression Domain

In order to look for enriched categories of transcription factors within the set of 363 cleaned YFP+/− DEGs (Fig. 2C), we used the online Transcription Factor Enrichment Calculator (https://dgrinevich.shinyapps.io/ShinyTF). Members of the AB13/VP1 transcription factor family that includes the REM and NGA family transcription factors were found to be statistically overrepresented (Supplemental Table S7; corrected P < 9.97E‐06). The REMs belong to the plant-specific B3 superfamily of transcription factors, and the expression of many REM family members is observed in meristematic tissues, such as the inflorescence meristem, floral meristem, and CMM (Franco‐Zorrilla et al., 2002; Swaminathan et al., 2008; Romanel et al., 2009; Wynn et al., 2011; Mantegazza et al., 2014a, 2014b). The numerical designations used to describe the REM family members in this article are taken from Romanel et al. (2009). In our study, six REM members were among the 363 statistically significant YFP+/− DEGs; five were found to have enriched expression in the YFP-positive sample, while one, REM25 (AT5G09780), was approximately 4-fold less abundant in the YFP-positive sample. The REM13 (AT3G46770) transcript level is enriched approximately 12-fold in the pSHP2-YFP-expressing cells. REM13 was predicted previously to be preferentially expressed in the inner integument, ovule primordia, and medial domain based on expression of transcriptomic data (Skinner and Gasser, 2009). We employed in situ hybridization to assay the expression pattern of the REM13 transcript during gynoecial development (Fig. 3). Using a REM13 antisense probe, we detected signal in the medial portions of the gynoecium corresponding to the CMM as early as stage 7. Expression also was observed in the initiating ovule primordia in stage 8 gynoecia and then continued to be detected in portions of the ovules at later developmental stages.

REM34/ATREM1 (AT4G31610; Franco-Zorrilla et al., 2002; Romanel et al., 2009), REM36 (AT4G31620; Mantegazza et al., 2014b), and VERDANDI (VDD/REM20; Matias-Hernandez et al., 2010; Mantegazza et al., 2014b) also displayed enriched expression levels in the YFP-positive sample of approximately 8-, 9-, and 6-fold, respectively. Published in situ hybridization patterns indicate enriched medial domain expression patterns for REM34/ATREM1 and VDD/REM20 (Franco-Zorrilla et al., 2002; Matias-Hernandez et al., 2010; Wynn et al., 2011). Additionally, the expression of AT5G60142, a previously unnamed member of the REM family, is enriched approximately 11-fold in the YFP-positive sample (Supplemental Table S3). AT5G60142 is an interesting candidate for functional studies that is located on chromosome V in tandem with REM1 (AT5G60140) and REM12 (AT5G60130) and shares a high degree of sequence similarity with these two genes as well as REM13 (Romanel et al., 2009; Mantegazza et al., 2014b). We propose to designate AT5G60142 as REM46.

Gene Set Enrichment Analysis

To gain global insights into underlying biological mechanisms of medial domain development and gynoecium development (Gremski et al., 2007; Crawford and Yanofsky, 2011). However, we chose to use the cleaned set of 363 YFP+/− DEGs for downstream analyses in order to reduce the likelihood of the inclusion of genes whose expression was significantly altered by the protoplasting process.
function, Gene Set Enrichment Analysis (GSEA) was performed for the 95 YFP-positive enriched DEGs in the medial domain. This analysis identified 147 Gene Ontology (GO) terms that were statistically overrepresented \((P < 0.01)\), including gynoecium development (GO:0048467), flower development (GO:0009908), response to gibberellin (GO:0009739), and auxin homeostasis (GO:0010252; Fig. 4; Supplemental Table S6). This GSEA analysis further suggests that the set of 95 genes enriched in the YFP-positive sample function as regulators of medial domain development.

In contrast, when performing GSEA with DEGs identified between the all-sorted/nonsorted samples, a different set of 304 overrepresented GO terms were identified, including response to stress (GO:0006950) and response to wounding (GO:0009611), suggesting that many of the genes identified as differentially expressed between the all-sorted/nonsorted samples reflect stress-induced changes in gene expression during protoplast/FACS sorting.

The Transcriptomic Signature of the SHP2-Expressing Cell Population Shares Commonalities with the Transcriptional Signatures of Other Meristematic Samples

In order to gain insight into the characteristics of the 363 YFP+/− DEGs identified from the SHP2 expression domain, we compared the expression profile of this set of genes across several different tissues. Using Spearman rank correlation analysis, we compared our data set with existing Arabidopsis RNA-seq transcriptomic data sets from whole flowers (Mizzotti et al., 2014) and aerial seedling tissues (Gene Expression Omnibus [GEO] accession no. GSE54125) as well as from laser-capture microdissected inflorescence meristems, floral meristems, and stage 3 flowers (Mantegazza et al., 2014a).

In the sample-wise hierarchical clustering (Fig. 5A), the transcriptomic profiles from the SHP2-expressing (YFP-positive) sample clustered more closely with the meristematic samples, while the YFP-negative and all-sorted samples clustered more closely with the whole-flower and whole-seedling samples. This suggests that the expression signature of the YFP-positive sample is more similar to that of the floral and inflorescence meristems and young flowers than it is to whole flowers or young vegetative seedlings (Fig. 5A).

Further supporting the similarity of the SHP2-expressing domain to other meristematic samples, the expression levels of \(GA20OX1\) (AT4G25420) and \(GA20OX2\) (AT5G51810) were both significantly depleted in the YFP-positive sample relative to the YFP-negative sample (Supplemental Table S10). \(GA20OX1\) and \(GA20OX2\) encode key biosynthetic enzymes of the plant hormone GA (Phillips et al., 1995). Levels of expression of \(GA20OX1\) and \(GA20OX2\) are low in the SAM relative to expression in the juxtaposed young organ primordia, and high levels of GA synthesis interfere with the maintenance of meristematic fate in the SAM (Hay et al., 2002; Jasinski et al., 2005). These data suggest that low levels of GA also may be associated with the meristematic nature of the CMM. Although not discussed here, the expression values of genes annotated with a role in ethylene signaling are found in Supplemental Table S10.

We additionally compared the medial domain transcriptional signature with data sets generated with the Affymetrix ATH1 array, allowing comparisons with transcriptomic signatures of a variety of cell types, including vascular and meristematic cell types from the...
Arabidopsis SAM isolated via FACS (Yadav et al., 2009, 2014). When these additional samples are included, the hierarchical clustering dendrogram (Fig. 5B) shows that the YFP-positive sample is more similar to the SAM cell types than to the vascular procambium (AtHB8) and phloem cell types (S17). This again suggests the meristematic character of the YFP-positive sample (Fig. 5B). One should be cautious, however, to interpret the results of this (or any) cross-platform (array/RNA-seq) comparison until validated cross-platform comparison methods are available. To the best of our knowledge, there is no clear consensus in the literature of a standard cross-platform comparison practice (Mudge et al., 2008; Bradford et al., 2010; Guida et al., 2011; Nookaew et al., 2012). Indeed, many researchers have used both platforms (array/RNA-seq) in the same experiment and compare the final results rather than find a way to directly compare the two technologies (Marioni et al., 2008; Nookaew et al., 2012; Xu et al., 2013; Wang et al., 2014; Zhao et al., 2014). Here, we employ a Spearman rank correlation, as it is less sensitive than the Pearson correlation to strong outliers, makes no assumptions about data distribution, and does not inflate type I error rates. This approach fits well with the data in this work, as samples do not cluster based on technology platforms but rather based on the apparent cell type similarities of gene reads per kilobase of transcript per million mapped reads (RPKM) expression levels.

Transcriptomic Analysis of the SHP2 Expression Domain Complements Existing Medial Domain and CMM Data Sets

Wynn et al. (2011) previously carried out a related transcriptomic study and identified many genes that were shown via in situ hybridization to be expressed preferentially in the developing medial domain of the wild-type gynoecium. When comparing the 95 enriched DEGs from our RNA-seq experiment (Supplemental Table S2; Supplemental Fig. S3) with a set of 210 medial domain-enriched genes from Wynn et al. (2011), 23 genes were found in common (Table III). The 24% overlap of these two gene sets is significantly higher than expected by chance (hypergeometric test, $P = 3.15 \times 10^{-30}$; Halbritter et al., 2012). Members of the REM, HEC, and NGA gene families, as well as several auxin homeostasis-related genes, were among the set of 23 genes identified in both experiments (Table III).
Reyes-Olalde et al. (2013) recently performed a comprehensive literature survey of genes that function during CMM development. They reported 86 protein-coding genes corresponding to transcription factors, hormonal pathways, transcriptional coregulators, and others of widely diverse functions. Fifteen of these 86 CMM developmental regulators are found within the set of 363 YFP-positive DEGs displaying a greater than 4-fold expression difference (Fig. 6, gene names marked containing)). Additionally, there are 32 DEGs that are statistically significant (at FDR < 0.01), where the fold enrichment or depletion is less than 4-fold. As many biologically significant DEGs may display less than 4-fold expression difference, these genes also are indicated (Fig. 6, marked with triple asterisks). Thus, a total of 47 of the 86 genes reported by Reyes-Olalde et al. (2013) also are detected as differentially expressed between YFP-positive and YFP-negative samples in our analysis. The expression profiles of the 86 genes reported by Reyes-Olalde et al. (2013) within the medial domain-enriched data set from this work, as well as within data from floral meristem-enriched samples (Mantegazza et al., 2014a), are displayed in the heat map in Figure 6 (RPKM values can be found in Supplemental Table S9).

### Transcript Isoforms in the Arabidopsis Medial Domain

One utility of transcriptome analysis through RNA-seq is the identification of novel alternative spliced transcripts, alternative transcription start sites (TSSs), and instances of isoform switching (Sims et al., 2014). To further characterize the transcriptome of the SHP2 expression domain at the isoform level, we first selected isoforms that showed a significant (α < 0.01) change in their expression between YFP+/− samples using Cufflinks/Cuffdiff. For this analysis, we did not apply a fold magnitude cutoff, thus capturing all isoforms with α < 0.01. To avoid transcripts that were affected by the cell-sorting procedure, we removed all isoforms that showed a significant (α < 0.01) expression level change between all-sorted/nonsorted samples. This resulted in

---

### Table II. List of genes displaying isoform-switching behavior between the SHP2-positive and SHP2-negative cell populations

<table>
<thead>
<tr>
<th>Isoform Identifier</th>
<th>TSS Group Identifier</th>
<th>CC</th>
<th>Nearest Reference Identifier</th>
<th>Isoform Relative RPKM for YFP POS</th>
<th>Isoform Relative RPKM for YFP NEG</th>
<th>Isoform Relative RPKM for ALL SORT</th>
<th>Isoform Relative RPKM for NO SORT</th>
</tr>
</thead>
<tbody>
<tr>
<td>TCONS_000000995</td>
<td>TSS74</td>
<td>=</td>
<td>AT1G02110.1 (DUF630)</td>
<td>11.54</td>
<td>37.20</td>
<td>15.31</td>
<td>19.62</td>
</tr>
<tr>
<td>TCONS_00007286</td>
<td>TSS5864</td>
<td>=</td>
<td>AT1G23340.1 (DUF239)</td>
<td>5.46</td>
<td>1.34</td>
<td>6.93</td>
<td>5.77</td>
</tr>
<tr>
<td>TCONS_00014129</td>
<td>TSS11387</td>
<td>=</td>
<td>AT2G42890.2 (MEI2-LIKE2)</td>
<td>7.90</td>
<td>2.20</td>
<td>5.42</td>
<td>5.77</td>
</tr>
<tr>
<td>TCONS_00041486</td>
<td>TSS33793</td>
<td>=</td>
<td>AT5G35050.3 (α/β-hydroxylase)</td>
<td>3.23</td>
<td>0.48</td>
<td>1.31</td>
<td>2.46</td>
</tr>
<tr>
<td>TCONS_00002986</td>
<td>TSS2424</td>
<td>=</td>
<td>AT1G43850.1 (SEU)</td>
<td>6.26</td>
<td>18.06</td>
<td>7.02</td>
<td>11.62</td>
</tr>
<tr>
<td>TCONS_00032425</td>
<td>TSS26386</td>
<td>=</td>
<td>AT4G32250.2 (protein kinase)</td>
<td>2.19</td>
<td>11.19</td>
<td>3.58</td>
<td>5.11</td>
</tr>
<tr>
<td>TCONS_00000744</td>
<td>TSS563</td>
<td>=</td>
<td>AT1G10570.2 (OTS2)</td>
<td>14.70</td>
<td>6.40</td>
<td>8.02</td>
<td>9.45</td>
</tr>
<tr>
<td>TCONS_00004152</td>
<td>TSS3371</td>
<td>=</td>
<td>AT1G61820.1 (BGLU46)</td>
<td>9.49</td>
<td>0.82</td>
<td>4.71</td>
<td>3.79</td>
</tr>
<tr>
<td>TCONS_00006607</td>
<td>TSS5332</td>
<td>=</td>
<td>AT1G14170.1 (KH domain containing)</td>
<td>10.18</td>
<td>2.62</td>
<td>5.68</td>
<td>5.39</td>
</tr>
<tr>
<td>TCONS_00014122</td>
<td>TSS11382</td>
<td>=</td>
<td>AT2G42830.2 (SHP2)</td>
<td>0.63</td>
<td>4.62</td>
<td>1.20</td>
<td>0.78</td>
</tr>
<tr>
<td>TCONS_00001240</td>
<td>TSS954</td>
<td>j</td>
<td>AT1G16710.2 (HAC12)</td>
<td>3.50</td>
<td>5.88</td>
<td>7.11</td>
<td>5.77</td>
</tr>
<tr>
<td>TCONS_00006659</td>
<td>TSS5370</td>
<td>=</td>
<td>AT1G4690.2 (MAP65-7)</td>
<td>39.97</td>
<td>16.72</td>
<td>28.41</td>
<td>18.98</td>
</tr>
<tr>
<td>TCONS_00039135</td>
<td>TSS31808</td>
<td>=</td>
<td>AT5G19130.1 (GPI transamidase)</td>
<td>3.46</td>
<td>17.27</td>
<td>13.34</td>
<td>12.45</td>
</tr>
<tr>
<td>TCONS_00033470</td>
<td>TSS27205</td>
<td>=</td>
<td>AT5G06610.1 (DUF620)</td>
<td>7.57</td>
<td>1.18</td>
<td>3.31</td>
<td>4.72</td>
</tr>
<tr>
<td>TCONS_00031775</td>
<td>TSS25869</td>
<td>=</td>
<td>AT4G23660.1 (PPT1)</td>
<td>2.56</td>
<td>9.59</td>
<td>3.23</td>
<td>6.70</td>
</tr>
</tbody>
</table>
4,555 YFP+/- differentially expressed isoforms (Supplemental Table S8). Within this set of isoforms differentially expressed between the YFP+/- samples, we sought to highlight multiple-isoform genes that showed major changes in the relative frequency of individual isoforms between the YFP-positive and YFP-negative samples. To this end, we estimated the relative frequency of each isoform as a percentage of the total expression for the gene. Among the 4,555 significantly differentially expressed isoforms, only 52 isoforms from multiple-isoform genes displayed changes of 20% or more in their relative frequency. The major isoform (most highly expressed isoform) differed between YFP+/- samples for only 15 genes (Table II).

Remarkably, the transcriptional coregulator SEUSS (SEU; AT1G43850), previously implicated in medial domain development (Franks et al., 2002; Azhakanandam et al., 2008), showed a significant increase of isoform AT1G43850.1 in the YFP-positive samples, while its second isoform, AT1G43850.2, did not change significantly between samples. As a result, isoform 1 was the major (predominant) isoform in YFP-positive cells and isoform 2 was the major (predominant) isoform in the other samples. The functional significance, if any, of this isoform switching is currently unknown. The two SEU mRNA isoforms differ only in the length of their 5' untranslated regions (UTRs; TAIR10). However, this difference in the 5' UTR raises the possibility that the SEU isoforms might be differentially regulated at a posttranscriptional level.

The SHP2 (AT2G42830) gene also was found in this list of 15 genes displaying an isoform switch (Table I). The two SHP2 isoforms differ in the splice donor site at the 5' end of intron 4 (TAIR10). This alternative splicing event is expected to add six nucleotides (two amino acids) to the AT2G42830.2 isoform. To our knowledge, the functional consequences of this alteration to the SHP2 protein sequence has not yet been investigated. The shorter AT2G42830.1 isoform is detected as the majority isoform in the YFP-positive samples.

The regulation of gene expression through alternative promoter usage or the use of alternative TSSs is observed frequently in multicellular organisms (Ayoubi and Van De Ven 1996). Using the same pipeline and criteria we employed to select differentially expressed isoforms in the YFP+/- samples, we identified 93 isoforms that were differentially expressed as a result of the use of alternative promoter/TSSs (Supplemental Table S8, promoter usage tab). Interestingly, one such promoter/TSS switch was found for the REVERSIBLY

Figure 6. Heat map representation of the expression profiles of previously identified regulators of CMM development. Expression profiles in RPKM are for the 86 genes reported by Reyes-Olalde et al. (2013) with functional roles during CMM development. Transcriptional profiles from this study (YFP POS = YFP positive, YFP NEG = YFP negative, ALL SORT = all sorted, and NO SORT = nonsorted) as well as from Mantegazza et al. (2014a) corresponding to flower stage 3 (FL. STAGE 3), floral meristem (FL. MERISTEM), and inflorescence meristem (IN. MERISTEM) are included. Genes color coded in red are those identified as DEGs between YFP-positive and YFP-negative samples (fold change > 4 and FDR < 0.001), while genes that displayed a statistically significant expression level (FDR < 0.01) between YFP-positive and YFP-negative samples (regardless of their fold change) are indicated with triple asterisks.
GLYCOSYLATED POLYPEPTIDE5 (RGP5) gene (isoform). Members of the RGP family (RGP1 and RGP2) involved in sugar metabolism are expressed in other Arabidopsis meristematic tissues, such as the root tip and the apical meristem of young seedlings (Drakakaki et al., 2006). In our work, the transcript level of RGP5 isoform 2 (AT5G16510.2) in the YFP-positive sample is 61% higher relative to the level of this isoform in the YFP-negative sample, while the level of isoform 1 (AT5G16510.1) is 75% lower (Fig. 7; Supplemental Table S8).

Table III. Overlapping DEGs between the 95 DEGs from this study that display enrichment in SHP2-expressing cells and 210 DEGs from Wynn et al. (2011) displaying reduced expression in the seu ant double mutant relative to other genotypes

<table>
<thead>
<tr>
<th>Annotation/Gene Symbol</th>
<th>Gene Identifier</th>
<th>RNA-seq (This Study)</th>
<th>Array (Log2 Scaled Expression Values from Wynn et al. [2011])</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Log2 Fold Change</td>
<td>Fold Change</td>
<td>P</td>
</tr>
<tr>
<td>HEC1</td>
<td>AT5G67060</td>
<td>5.9</td>
<td>59.4</td>
</tr>
<tr>
<td>Cys proteinase</td>
<td>AT1G03720</td>
<td>4.1</td>
<td>17.7</td>
</tr>
<tr>
<td>REM13</td>
<td>AT3G46770</td>
<td>3.6</td>
<td>12.2</td>
</tr>
<tr>
<td>ATCEL2</td>
<td>AT1G02800</td>
<td>3.4</td>
<td>10.4</td>
</tr>
<tr>
<td>LAX1</td>
<td>AT5G01240</td>
<td>3.3</td>
<td>9.7</td>
</tr>
<tr>
<td>LTP2</td>
<td>AT2G38530</td>
<td>3.3</td>
<td>9.5</td>
</tr>
<tr>
<td>AMP-dependent synthetase</td>
<td>AT1G21540</td>
<td>3.2</td>
<td>9.3</td>
</tr>
<tr>
<td>ACO5</td>
<td>AT2G27880</td>
<td>3.0</td>
<td>8.0</td>
</tr>
<tr>
<td>UGT84A2</td>
<td>AT3G21560</td>
<td>2.9</td>
<td>7.6</td>
</tr>
<tr>
<td>TAA1</td>
<td>AT1G070560</td>
<td>2.8</td>
<td>6.9</td>
</tr>
<tr>
<td>ATDDOF5.8</td>
<td>AT5G66940</td>
<td>2.7</td>
<td>6.6</td>
</tr>
<tr>
<td>VODD</td>
<td>AT5G18000</td>
<td>2.6</td>
<td>6.0</td>
</tr>
<tr>
<td>MCT1</td>
<td>AT1G37140</td>
<td>2.6</td>
<td>5.9</td>
</tr>
<tr>
<td>Unknown</td>
<td>AT2G41990</td>
<td>2.5</td>
<td>5.6</td>
</tr>
<tr>
<td>Protein kinase superfamily protein</td>
<td>AT1G74490</td>
<td>2.5</td>
<td>5.5</td>
</tr>
<tr>
<td>NGA3</td>
<td>AT1G01030</td>
<td>2.3</td>
<td>5.1</td>
</tr>
<tr>
<td>NCA1</td>
<td>AT2G46870</td>
<td>2.3</td>
<td>4.8</td>
</tr>
<tr>
<td>Cystatin/monellin</td>
<td>AT1G03710</td>
<td>2.3</td>
<td>4.8</td>
</tr>
<tr>
<td>YUC4</td>
<td>AT5G11320</td>
<td>2.3</td>
<td>4.8</td>
</tr>
<tr>
<td>Heavy metal transport</td>
<td>AT1G56210</td>
<td>2.2</td>
<td>4.7</td>
</tr>
<tr>
<td>ANAC098</td>
<td>AT5G53950</td>
<td>2.1</td>
<td>4.2</td>
</tr>
</tbody>
</table>

Figure 7. Differential expression of RGP5 isoforms. A promoter/TSS switch was found for the RGP5 gene (AT5G16510). Isoform 2 (AT5G16510.2) in the YFP-positive sample is 61% higher relative to the level of this isoform in the YFP-negative sample, while the level of isoform 1 (AT5G16510.1) is 75% lower (Fig. 7; Supplemental Table S8).

Auxin Homeostasis and the Development of the Gynoecial Medial Domain

Auxins are a class of plant hormones that regulate growth and development (Woodward and Bartel, 2005; Sauer et al., 2013). The most common plant auxin is indole-3-acetic acid. The regulation of auxin homeostasis (including synthesis, response, transport, inactivation, and degradation) plays an essential role in patterning the gynoecium and other lateral organs (Woodward and Bartel, 2005; Sehra and Franks, 2015).
The role of auxin during the development of the medial and lateral domains of the gynoecium is less clearly defined; however, recent studies suggest that auxin homeostasis mechanisms are likely to be distinct in medial and lateral domains (Larsson et al., 2014; Moubayidin and Ostergaard, 2014; Sehra and Franks, 2015).

To better analyze auxin homeostatic mechanisms during medial domain development, we examined the expression of 127 genes with an annotated function in auxin homeostasis. Of these 127 genes, 80 were expressed in our data set and 60 were differentially expressed at FDR $< 0.01$ in the YFP-positive sample, without applying a fold enrichment filter (Supplemental Table S10). The expression of TRYPTOPHAN AMINOTRANSFERASE OF ARABIDOPSIS1 (TAA1) and YUC4, two genes encoding proteins in the auxin synthetic pathway, were strongly enriched (more than 4-fold) in the YFP-positive samples, as predicted from previously published expression patterns indicating enriched expression within the medial portions of the gynoecium (Zhao et al., 2001; Cheng et al., 2006; Stepanova et al., 2008; Tao et al., 2008; Trigueros et al., 2009; Martínez-Fernández et al., 2014). Within the PINFORMED (PIN) family of polar auxin transporters, the expression levels of PIN1, PIN3, and PIN7 were enriched significantly in the YFP-positive sample (Supplemental Table S10). This is consistent with the reported expression patterns at the protein level of these PIN transporters within the medial domain of the gynoecium (Benková et al., 2003; Blilou et al., 2005; Larsson et al., 2014; Moubayidin and Ostergaard, 2014).

Auxin induces gene expression through a family of transcription factors called AUXIN RESPONSE FACTORS (ARFs; Woodward and Bartel, 2005). At a fold change level of 1.5 and FDR $< 0.01$, 10 ARFs were enriched in the YFP-positive sample (ARF1, ARF2, ARF3, ARF4, ARF5, ARF6, ARF7, ARF8, ARF16, and ARF18), while no ARFs were identified as depleted in the YFP-positive sample (Supplemental Table S10). Our data suggest that

---

**Figure 8.** Differential expression of TAS3 and ARF genes. Expression of the ARFs (ARF2, ARF3, ARF4) and TAS3 transcripts is shown. Expression levels of ARF2, ARF3, and ARF4 are significantly enriched in the YFP-positive sample relative to the YFP-negative sample at FDR $< 0.01$. Expression levels of the TAS3 genes AT5G49615 (TAS3b) and AT3G17185 (TAS3), which negatively regulate the expression of ARF2, ARF3, and ARF4 expression (Williams et al., 2005), are significantly reduced (FDR $< 0.01$) in the YFP-positive sample. All error bars indicate confidence intervals as reported by Cufflinks. Three asterisks indicate statistically significant differences at FDR $< 0.01$; ns indicates not statistically different.
these ARF family members may be expressed preferentially in the medial domain and play a role during the development of this meristematic tissue. Previous studies have documented gynoecial developmental defects in arf3/ettin mutants (Sessions and Zambryski, 1995) as well as in arf6 arf8 double mutants (Nagpal et al., 2005; Wu et al., 2006). Interestingly, the levels of the precursor transcripts for two TRANS-ACTING SIRNA3 (TAS3) genes (AT5G49615 and AT3G17185) were reduced significantly (FDR, 0.01) in the YFP-positive sample (Fig. 8; Supplemental Table S8). The trans-acting small interfering RNAs that are encoded by the TAS3 genes negatively regulate the levels of ARF2, ARF3, and ARF4 transcripts (Williams et al., 2005). Thus, the enrichment of ARF2, ARF3, and ARF4 transcript levels in the SHP2 expression domain may be due in part to a reduction in the level of expression of the TAS3-encoded trans-acting small interfering RNAs in the medial domain.

SQUAMOSA PROMOTER-BINDING PROTEIN-LIKE3 and the cis-NAT Antisense Gene AT2G33815

The SQUAMOSA PROMOTER-BINDING PROTEIN-LIKE (SPL) genes function in the regulation of the transition from juvenile to adult growth phases and the regulation of shoot regenerative capacity (Wu and Poethig, 2006; Wang et al., 2008, 2009; Zhang et al., 2015). In our study, the expression of SPL3 (AT2G33810) was more than 4-fold lower in the YFP-positive sample relative to the YFP-negative sample. SPL3 encodes a DNA-binding protein directly regulating APETALA1 (AT1G69120), a key regulator of floral meristem identity specification (Leal Valentim et al., 2015). Interestingly, the expression of the cis-NAT antisense gene AT2G33815, complementary to portions of the SPL3 gene, also was reduced significantly approximately 4.5-fold in the YFP-positive sample (Cufflinks data in Supplemental Table S8). In both plant and animal systems, the coexpression of overlapping sense and antisense transcripts can lead to the generation of cis-NAT-derived small interfering RNAs that can regulate gene expression at transcriptional and posttranscriptional levels (Zhang et al., 2015). The expression of another regulator of SPL3 activity, miRNA157D (AT1G48742), also was reduced significantly in the YFP-positive samples. miRNA157D reduces the translation of the SPL3 transcript by acting through a miRNA156/157-responsive element in the SPL3 3′ UTR (Gandikota et al., 2007; Wang et al., 2009). These data suggest that the miRNA156/157/SPL module may act during medial domain development and may be regulated by the cis-NAT antisense gene AT2G33815. A complete list of differentially expressed natural antisense, transposable element, and other non-protein-coding transcripts identified as differentially expressed by Cufflinks, DESeq2, and edgeR is found in Supplemental Table S8.

Protoplasting-Induced Stress Genes

While the predominant focus of this work was to perform transcriptomic analysis in medial domain-enriched cells (YFP+/−), transcripts induced by the protoplasting and sorting process (all-sorted/nonsorted)
also were identified (see “Materials and Methods”). To facilitate the visualization of all samples, we generated an interactive six-way Venn diagram using the Web-based tool InteractiVenn (Heberle et al., 2015). By uploading Supplemental File S1 to InteractiVenn (http://www.interactivenn.net), mousing over, and clicking on the numbers in the Venn diagram, researchers will find gene identifiers from DEGs between YFP+/− and all-sorted/nonsorted samples (three programs and two comparisons). As expected, when comparing such different types of samples (all-sorted/nonsorted and YFP+/−), few DEGs (26) overlapped across the six samples (Fig. 9). The lack of overlap of DEGs across the entire experiment indicates that the YFP+/− DEGs reported here are not a result of protoplasting stress-induced processes.

When comparing the protoplasting-induced gene set from this work (all-sorted/nonsorted DEGs) with those induced due to FACS-sorting methodology in SAM by Yadav et al. (2009) and in roots as reported by Birnbaum et al. (2005), few DEGs were found in common (seven). Yadav et al. (2009) and in roots as reported by Birnbaum et al. (2005). Few DEGs (26) overlapped across the six samples (Fig. 9). The lack of overlap of DEGs across the entire experiment indicates that the YFP+/− DEGs reported here are not a result of protoplasting stress-induced processes.

CONCLUSION

Despite the importance of the gynoecial medial domain in ovule development, no domain-specific transcriptome has been reported previously, mainly due to the difficulty of isolating the meristematic cells from which ovules are derived. In this work, we developed a novel FACS-based system using the SHP2 expression domain using a GAL4/pUAS-based two-component system that, when combined with flower synchronization and flow cytometry, allowed for the efficient isolation of medial domain cells expressing SHP2. The quality and quantity of biological samples that can be recovered with our system enables cell type- and strand-specific RNA-seq transcriptomic analysis and opens up possibilities for small RNA, metabolomic, and proteomic analyses (Petersson et al., 2009; Breakfield et al., 2012; Petricka et al., 2012; Li et al., 2013; Moussaieff et al., 2013). This approach, coupled with high-throughput RNA-seq, has yielded a unique and novel snapshot of the gynoecial medial domain transcriptome and a set of candidate regulators of medial domain development for future functional analysis.

MATERIALS AND METHODS

Construction of psHP2-GAL4; pUAS-3xYpet

Dual-Construct Lines

The SHP2 promoter fragment was amplified from Columbia-0 wild-type genomic DNA using the primers proSHP2gwF1 (5′-CACCATCTCCAAGGATCTGTAGC-3′) and proSHP2gwR1 (5′-CATTCTTATAACGGCTTCACTGAAAG-3′). This fragment contains the sequences from −2,170 to +1 relative to the SHP2 ATG and includes the 5′ UTR, the first intron, and the first Met codon of SHP2. This promoter was shown previously to mimic the endogenous SHP2 expression pattern (Columbia-0 and Landsberg erecta). This genomic fragment was cloned into the pENTR/D-TOPO vector (invitrogen) to create plasmid LJM01 and then shuttled via Gateway LR reaction (Invitrogen) into the destination vector JMA589 (i.e., pEarleygate303-GAL4) to create plasmids AAS003. Transgenic Arabidopsis (Arabidopsis thaliana) lines were created by Agrobacterium tumefaciens-mediated transformation of the AAS003 plasmid into the S. No. 1880 seed stock that contained the pGW2B-pUAS-3xYpet responder construct (see below), generating the pSHP2-GAL4; pUAS-3xYpet dual-construct line (S. No. 1896), referred to as pSHP2-YFP.

The pSHP2-YFP plants were crossed to the ap1-1 Wellmer floral induction system (Wellmer et al., 2006) as described below.

JMA589 (pEarleygate303-GAL4) was a modified pEarleygate303 (Earley et al., 2006) plasmid in which the reporter was replaced by the coding sequences from the GALA4 Saccharomyces cerevisiae transcriptional activator. To achieve this, pEarleygate303 was cut with Ncol (New England Biolabs) and SpeI (New England Biolabs). Then, fusion PCR was used to create the insert that fused the GAL4 sequences to the deleted portions of pEarleygate303. This required three PCRs: first PCR with primers pEarl303NcolFor (5′-TGCCCAATAATGGCAATCCTCT-3′) and pEarl303Rev_GAL4 (5′-ATGGAGGACAGGAGCTTCATACACAGATCTTCTTCAGAGA-3′); second PCR with primers GAL4revEarl303 (5′-TCTCTGGAAGAACGTGTTGATGA-GCCTCTCTCCTCAT-3′) and GAL4Rev (5′-CCGGACTAGTCTACCCCGTACTCGTCAAA-3′); and then a fusion PCR joining these two fragments using the external primers to amplify. The product of the fusion PCR was double digested with Ncol/SpeI and ligated into Ncol/SpeI-cat pEarleygate303.

JMA382 (pUAS-pGW2B) was created from pGW2B (Nagakawa et al., 2007) by replacing the p3SS sequences in pGW2B with pUAS sequences (HindII/XbaI sites used). A Gateway LR reaction was then used to move the 3xYpet cassette from JMA710 (pENTR/D-TOPO-3xYpet) into JMA382, creating vector JMA721 (i.e. pGW2B-pUAS-3xYpet). Homozygous single insertion site transgenic lines harboring JMA721 were then generated (S. No. 1800).

Plant Material

In a wild-type inflorescence, cells expressing SHP2 represent a small percentage of the total cells. Additionally, wild-type inflorescence contains a full range of developmental series of floral stages. The Wellmer floral synchronization system (Wellmer et al., 2006) was used to maximize the amount of gynoecial tissue from floral stages 6 to 8 (Smyth et al., 1990). The Wellmer group kindly provided pAP1-AP1:GR; ap1; cal seeds (kanamycin resistant in the Landsberg erecta background; S. No. 1927). The pSHP2-GAL4; pUAS-3xYpet dual-construct plants (S. No. 1896) were crossed to pAP1-AP1:GR; ap1; cal. Lines homozygous for erecta, ap1, cal, and the transgenes were selected in F2 and F3 generations (generating S. No. 2060). Because of the mixed ecotype cross (Columbia-0 and Landsberg erecta), lines that were erecta homozygous mutant and gave a consistent YFP expression pattern and consistent inducibility of the AP1-GR activity were selected before the generation of protoplasts. Plants were grown under constant light and temperature at 22°C to minimize circadian transcriptional fluctuations. To induce flowering in the transgenic plants, 20 μm of the synthetic steroid hormone dexamethasone (Sigma) in 0.015% (v/v) Silwet L77 was applied directly (spray application) approximately 30 d after planting (Wellmer et al., 2006). Inflorescences were collected for protoplast generation approximately 120 h after dexamethasone-induced floral synchronization. When collecting samples for protoplast preparation, five to six inflorescence heads were fixed for chloral hydrate clearing and DIC microscopy to determine the developmental stages of the flowers of the inflorescence samples. Additionally, before protoplasting, whole inflorescences were collected and frozen immediately in liquid nitrogen for analysis of the transcriptional starting state of the nonprotoplasted tissue (nonsorted samples; see “Experimental Design”).

Experimental Design

Material for RNA samples was gathered from batches of plants grown at 1-week intervals to generate biological replicates (material from each week was considered as a biological replicate). To reduce variability between biological replicates due to environmental heterogeneity within the growth chamber, each bio-replicate was drawn from a pool that contained plants grown within three different chamber positions. Four biological replicates of each of four tissue samples (YFP-positive, YFP-negative, all-sorted, and nonsorted samples) were...
collected (16 samples total). Whole inflorescences were collected for nonsorted samples and immediately frozen in liquid nitrogen before RNA isolation (i.e. these samples were not subjected to protoplasting or FACS sorting). The all-sorted samples represented the total population of protoplasts that come off the FACS machine after debris and broken cells are removed based on sorting gates (Supplemental Fig. S1). The YFP-positive and YFP-negative protoplast populations are processed equivalently to the all-sorted samples except that a final FACS-sorting gate is used to divide the all-sorted protoplasts into YFP-positive and YFP-negative samples (Supplemental Fig. S1). RNA was isolated from these three protoplast populations as well as from entire nonprotoplasted inflorescences (nonsorted). The YFP-positive, YFP-negative, and all-sorted samples were prepared and collected as described below (see “Protoplast Recovery and Cell Sorting”).

Protoplast Recovery and Cell Sorting

Protoplasts from the S. No. 2060 plants were generated according to the protocol of Birnbaum et al. (2005), with adaptations for inflorescence plant material. Inflorescences (approximately 200) were hand collected with forceps and/or scissors and chopped with a Personna double-edge prep blade (American Safety Razor; 74-002) within a 15-min period. Cell wall polysaccharides were digested by immersing the chopped plant material in 10 mL of filter-sterilized solution B in a 50-mL Falcon tube. Solution B (prepared according to Birnbaum et al., 2005) is prepared from solution A (10 mM KCl, 2 mM MgCl2, 0.2 mM MES, and 600 mM mannitol) to which cell wall-digesting enzymes were added (final concentrations of 1.5% (w/v) Cellulase [Yakult], 1% (w/v) Pectolyase [Yakult], and 1% (w/v) Hemicellulase [Sigma]). This mixture is then dissolved by gentle swirling, covered in foil, and warmed in a water bath at 55°C for 10 min to inactivateDNases and proteases. After cooling to room temperature, CaCl2 (2 mM final) and bovine serum albumin (0.1% (w/v) final) were added, and the solution was filter sterilized through a 25-μm filter.

After 1 h of incubation at room temperature with occasional gentle agitation, 10 mL of the protoplast-rich solution B was filtered through a 70-μm filter basket to a 50-mL Falcon tube. A 10-mL rinse of solution A was applied directly to the material left in the 70-μm filter basket to rinse through any protoplasts left behind. Protoplasts were spun at 500 g at 10°C for 10 min; the majority of the supernatant was removed by aspiration, being careful not to disturb the protoplast pellet, which is typically not tightly compacted. Protoplasts were resuspended in 25 mL of solution A as a rinse step to remove wall-digesting enzymes. Protoplasts were filtered again through a 50-μm filter mesh to a new tube, adding 5 mL of solution A to again rinse through any protoplasts stuck in the filter. Protoplasts were spun again at 500 g for 10 min. The majority of the supernatant was removed, leaving 2 mL of the protoplasts in solution after the second centrifugation step. Propidium iodide (5 μg mL⁻¹ final concentration) was added to the protoplasts (to allow separation of broken protoplasts), and a final filtering step through a 30-μm mesh filter (CellTrics; Partec) was carried out before loading onto the FACS machine.

Flow cytometry through FACS sorting (Moflo XDP; Beckman Coulter) was used to isolate the YFP-expressing cells from the total pool of cells. The FACS machine was equipped with a cooling device (set to 10°C) and occasionally agitated during the approximately 40 min of centrifugation that was performed to remove trace amounts of ethanol. Total RNA was then eluted with 10 μL of RNase-free water. A second elution was performed with another 10 μL of RNase-free water. It is worth noting that one biological replicate (fourth biological replicates) from the YFP-positive protoplasts was lost at this point, leaving only biological replicates for this tissue sample and yielding a total of 15 samples sequenced in two lanes and used for the experiment.

Prior to high-throughput sequencing, qRT-PCR was conducted on YFP-positive and YFP-negative samples using the 2−ΔΔct method as suggested by Schmittgen and Livak (2008) to assess the relative gene expression of the specific medial domain markers SHP2 and NGA1. Total isolated RNA was quantified using fluorometric quantitation (Qubit RNA Assay Kit; Life Technologies) for both YFP-positive and YFP-negative samples (approximately 100 ng). The SuperScript III First-Strand Synthesis System (Invitrogen/Life Technologies) was used to generate complementary DNA (cDNA; diluted 1:4 prior to qRT-PCR analysis) from total RNA. The qRT-PCR assay was performed (Thermal Cyclers from Applied Biosystems) using a SYBR Green mix (QuantiTect SYBR Green PCR Kits; Qiagen). Three biological replicates of the YFP-positive and YFP-negative samples were included, and each biological replicate was assayed in triplicate. The expression levels of the ADENINE PHOSPHOCHORYL TRANSFERASE1 (ATIG24750) gene was used for normalization.

Bar Plots

Bar plot graphs were constructed using the R packages bear (http://sourceforge.net/projects/jylee-r-packages/files/bear/) and plyr (https://cran.r-project.org/web/packages/plv/index.html) to calculate means, s.d., and confidence intervals and ggplot2 (https://cran.r-project.org/web/packages/ggplot2/index.html) to generate the plots.

Library Preparation and mRNA Sequencing

Total RNA isolated was quantified using fluorometric quantitation (Qubit RNA Assay Kit; Life Technologies), and RNA quality was assessed using the Agilent 2100 Bioanalyzer. The RNA integrity number for the 15 samples was higher than 7.3, which is above the Illumina threshold for library construction (greater than 7). Strand-specific cDNA libraries were constructed from approximately 100 ng of total RNA using a NEB Ultra Directional Library Prep Kit for Illumina (New England Biolabs). The average size of the CDNA fragments was approximately 250 bp. The 15 bar-coded libraries were pooled, and single-end sequencing was performed in a HiSeq 2500 device (Illumina) with HiSeq SR Cluster Kit version 4 for the flow cell and HiSeq SBS version 4 for sequencing reagents. cDNA libraries were sequenced in 125 cycles plus seven cycles for multiplexed samples. Sequencing was performed in two lanes of a flow cell; all 15 libraries were sequenced twice, and the results from the two independent
lanes were analyzed as technical replicates. As no lane-specific effects were observed during data analysis, the reads from each lane were pooled for the analysis of DEGs (see "Table Counts and Technical Replicates").

Bioinformatics Analysis

All bioinformatics analyses were performed on a server cluster with 128 gigabytes of RAM, two 8-core processors per compute node, and a Ubuntu Linux-Distribution 12.04 operating system using the Simple Linux Utility Resource Management queue management system at the Bioinformatics Research Center at North Carolina State University.

Read Processing

Quality control and preprocessing of metagenomic data were performed using FastQC software (Schmieder and Edwards, 2011). Adapters and low-quality sequences were filtered out with Ea-Utils software (Lindgreen, 2012). Reads with phred-like quality score greater than 30 and read length greater than 50 bp were kept and aligned against the TAIR 10 reference genome.

Sequence Alignment to the Arabidopsis Genome

The splice junction mapper TopHat2 (version 2.0.10; Trapnell et al., 2009) was used to align filtered RNA-seq reads to the TAIR 10 genome (Ensembl annotation) downloaded from the iGenome database (http://support.illumina.com/sequencing/sequencing_software/igenome.html). Default parameters for TopHat2 were used except for strand specificity (library-type = fr-firststrand) to match to the first strand of cDNA synthesized (antisense to the mRNA) and maximal intron length (~1 000), as it has been shown that the large majority of known introns are smaller than the selected threshold (Li et al., 2013). To align reads solely and exclusively against TAIR 10 annotated gene models, the arguments –T (transcriptome only) and –no-novel-juncs (no novel junction) also were included. Unique mapped reads were extracted from the TopHat2 output binary (BAM) file using samtools (Li et al., 2009) and selecting for the NH:i:1 two-character string tag. Only uniquely mapped reads were used for downstream analysis.

Table Counts and Technical Replicates

The HTSeq: Analyzing High-Throughput Sequencing Data with Python software (Anders et al., 2015) was used with default parameters except for the stranded = reverse mode to generate tables-counts for downstream differential expression analysis for the R packages edgeR (Robinson et al., 2010) and DESeq2 (Love et al., 2014). Using edgeR, we assessed the gene level variance versus log gene expression level among technical replicates (corresponding to two lanes in the flow cell of the Illumina HiSeq2500). A linear-dependent Poisson distribution was observed for technical replicates (Supplemental Fig. S5), in accordance with several studies (Marioni et al., 2008; Anders and Huber, 2010; Robinson et al., 2010). Thus, differential gene expression analysis was performed using pooled technical replicates.

Gene Expression and Differential Gene Expression

Gene expression and differential gene expression analyses were carried out using the R packages edgeR (Robinson et al., 2010) and DESeq2 (Love et al., 2014) and the Linux-based Cueflinks program (version 2.2.1) -G option (Trapnell et al., 2012) for DEGs and transcripts. To facilitate future use of these data sets, all the expressed genes identified and their expression values (F/FPKM) in the YFP+/− (Supplemental Table S3) and all-sorted/ nonsorted (Supplemental Table S4) samples are included. Filters were applied to determine if a gene was detected, abiding by the suggestions of statisticians and bioinformaticians (Bourgon et al., 2010; Rau et al., 2013; Sonesson and Dolorenz, 2013; Love et al., 2014; Pimentel et al., 2014; Seyedinassari et al., 2015) as a means to enrich for true DEGs, to reduce type I error, and to improve P value adjustment. The edgeR function (Robinson et al., 2010) cpm (counts per million) was used to discard those genes whose cpm was lower than a threshold of two reads per gene in at least three biological replicates, as suggested in the edgeR vignette. For Cueflinks, a minimum RPKM of 5 was set for a gene to be expressed, following the criteria of Suzuki et al. (2015). According to Sims et al. (2014), 80% of genes can be accurately quantified with FPKM/TPM > 10. DESeq2 performs independent filtering using the results function, as described in the DESeq2 vignette (Love et al., 2014). An FDR cutoff of less than 0.01 was used to determine DEGs in all three programs. The Gene Regulatory Information Server was used to identify transcription families in the data set (Yilmaz et al., 2011). Enriched categories of transcription factors within the set of cleaned 363 YFP+/− DEGs was assessed with the online transcription factor enrichment calculator tool (https://dgrinevich.shinyapps.io/ShinyTF).

Venn Diagrams and Heat Maps

Venn diagrams were constructed using the R package VennDiagram (Chen and Boutot, 2011) and the Web-based tool package InterActiVenn (Heberle et al., 2015). Heat maps were produced using the R package pheatmap (http://cran.r-project.org/web/packages/pheatmap/index.html). RPKM normalization by gene length and library size values was produced using the rpkm function from edgeR (Robinson et al., 2010). To calculate gene length, a TAIR 10 gene length list (coding sequence plus UTRs) was constructed by extracting length information from the TAIR 10 GFF file with homemade Perl script. Genes with multiple isoforms were collapsed, and length was calculated using the longest one. RPKM values were then calculated for clustering purposes and to have an intermediate point of comparison between Cueflinks, edgeR, and DESeq2. Samples were clustered (default clustering) with parameters provided in the software. The R package colorRamp (http://stat.ethz.ch/R-manual/R-devel/library/grDevices/html/colorRamp.html) was used to produce a gradient of color values corresponding to gene fold change values.

GSEA

GO enrichment tests were performed using the R package topGO (Alexa and Rahnenführer, 2009), with the classic algorithm (where each GO category is tested independently) and the Fisher statistical test for biological processes, molecular function, and cellular component. Enrichment analysis was performed separately for all the genes that were differentially expressed between the YFP+/− samples and the all-sorted/nonsorted samples. Network analysis of GO terms was performed using the BiNGO (Maere et al., 2005) plugin for Cytoscape (Shannon et al., 2003). GO terms for the 268 genes identified as depleted in the YFP-positive sample, as well cellular component and molecular function for the YFP+/− sample, can be found in Supplemental Table S6.

Dendrograms

The R Dst function was used to compute a distance matrix using the Spearman method (Spearman test rank correlation) and the R Cor function to compute the variance of the matrix. To perform hierarchical clustering, the hcust function in R was used. All statistical analyses were performed in R version 3.0.2. Dendrogram plots were built using the R ape package with edge.color = blue.

Confocal Microscopy

Confocal microscopy was performed using a Zeiss LSM 710 microscope model (Zeiss Axio Observer Z.1) with objective type Plan-Apochromat 20×/0.8 M27. Z-stack intervals were set to 2 μm, and the total thickness of the stack was 62 μm.

Chloral Hydrate Clearing and DIC Microscopy

Inflorescence samples were fixed in a solution of nine parts ethanol to one part acetic acid for 2 h at room temperature and then washed twice in 90% (v/v) ethanol for 30 min each wash. Inflorescences were transferred to Hoyer’s solution (70% (v/v) chloral hydrate, 4% (v/v) glycerol, and 5% (w/v) gum arabic) and allowed to clearing for several hours to overnight. Samples were then dissected in Hoyer’s solution. The dissected inflorescence heads were mounted in Hoyer’s solution under coverslips and examined with DIC optics on a Zeiss Axioskop 2 to determine the floral stages.

In Situ Hybridization

For in situ hybridization analysis, Arabidopsis Columbia-0 flowers were fixed and embedded in paraffin as described previously (Franks et al., 2002;
Wynn et al., 2011). Sections of plant tissue were probed with digoxigenin-labeled antisense and sense RNA probes (Roche). Probes corresponded to nucleotides +686 to +920 of REM13 relative to the transcriptional start site of the coding sequence using the following oligonucleotides to amplify the template: REM13ISH_FWD (5'-AAAAATGAGCCCGCATATGG-3') and REM13ISH_REV (5'-TCGTGACACCAAGCTGATA-3'). Hybridization and immunological detection were performed as described previously (Franks et al., 2012; Wynn et al., 2013).

Illumina sequencing raw data (fastq) have been submitted to the GEO database (accession no. GSE74458).

Supplemental Data

The following supplemental materials are available.

Supplemental Figure S1. Sorting gates used to select YFP samples and the resorting of the YFP-positive cells to assess sample purity.

Supplemental Figure S2. qRT-PCR enrichment of the medial domain genes SHP2 and NGA1 and the gene TUB.

Supplemental Figure S3. Expression profiles for the 363 DEGs (fold change > 4 and FDR < 0.001) across all four samples (YFP positive, YFP negative, all sorted, and nonsorted).

Supplemental Figure S4. Venn diagram comparison of stress-induced genes due to the protoplast/FACS-sorting procedure.

Supplemental Figure S5. Gene level variance versus log gene expression level among technical replicates.

Supplemental Table S1. Summary of RNA-seq data.

Supplemental Table S2. DEGs from the YFP+/− and all-sorted/nonsorted comparison.

Supplemental Table S3. All the expressed protein-coding genes identified with three different programs between the YFP+/− samples.

Supplemental Table S4. All the expressed protein-coding genes identified with three different programs between the all-sorted/nonsorted samples.

Supplemental Table S5. Raw high-throughput count data for the YFP+/− and all-sorted/nonsorted comparison.

Supplemental Table S6. GSEA for the YFP+/− and all-sorted/nonsorted comparison, including biological process, molecular function, and cellular component.

Supplemental Table S7. Transcription factor families identified in the DEGs from YFP+/− and their statistical enrichment.

Supplemental Table S8. Isoform expression, regulation of gene expression by alternative promoters, and antisense transcripts identified by Cufflinks, edgeR, and DESeq2.

Supplemental Table S9. Expression profile (RPKM) of the 86 genes described by Reyes-Olalde et al. (2013) expressed in the medial domain.

Supplemental Table S10. Hormone-related genes present in our data set.

Supplemental File S1. Data file to upload to the Web-based tool package InteractiVenn.

ACKNOWLEDGMENTS

We thank Sarah Schuetz (CVM, Flow Cytometry Facility, North Carolina State University [NCSU]) and the Genomic Sciences Laboratory Research Facility (NCSU) for library preparation and Illumina sequencing; William Thomson and Emily Wear (NCSU) for FACs-sorting assistance; Frank Wellmer and Diamand O’Maoileidigh (Smurfit Institute, Trinity College of Dublin) for the pAPl-AP2::GR; ap1; cif floral synchronization system; Colleen Doherty (NCSU) for help with Cytoscape/BiNGO analysis; Maria Angeles De Luis Balaguer (NCSU) for assistance with cross-platform expression correlation approaches; José Alonso and Ross Sozzani (NCSU) for thoughtful comments on the article; Jigar Desai, Dmitry Grinevich, and Colleen Doherty (NCSU) for help with the transcription factor enrichment analysis; Aureliano Bombarely (Virginia Tech) for help with GO term analysis; and Eva Johannes (CMIF, Molecular Imaging Facility, NCSU) for laser scanning confocal microscope assistance.

Received December 7, 2015; accepted March 14, 2016; published March 16, 2016.

LITERATURE CITED


Crawford BCW, Yanofsky MF (2011) HALF FILLED promotes reproductive tract development and fertilization efficiency in Arabidopsis thaliana. Development 138: 2999–3009


