1 **High-resolution Expression Profiling of Selected Gene Sets during Plant Immune**
2 **Activation**

3

4 Pingtao Ding[1, 2, 3, *, @], Bruno Pok Man Ngou[1, 3, @], Oliver J. Furzer[1, 4], Toshiyuki Sakai[1], Ram
5 Krishna Shrestha[1], Dan MacLean[1], Jonathan D. G. Jones[1, *, @]

6

7 **SUMMARY**

8 Sequence capture followed by next-generation sequencing has broad applications in cost-
9 effective exploration of biological processes at high resolution [1, 2]. Genome-wide RNA
10 sequencing (RNA-seq) over a time course can reveal the dynamics of differential gene
11 expression. However, in many cases, only a limited set of genes are of interest, and are
12 repeatedly used as markers for certain biological processes. Sequence capture can help
13 generate high-resolution quantitative datasets to assess changes in abundance of selected
14 genes. We previously used sequence capture to accelerate *Resistance* gene cloning [1, 3, 4],
15 investigate immune receptor gene diversity [5] and investigate pathogen diversity and
16 evolution [6, 7].
17 The plant immune system involves detection of pathogens via both cell-surface and
18 intracellular receptors. Both receptor classes can induce transcriptional reprogramming that
19 elevates disease resistance [8]. To assess differential gene expression during plant
20 immunity, we developed and deployed quantitative sequence capture (CAP-I). We designed
21 and synthesized biotinylated single-strand RNA bait libraries targeted to a subset of defense
22 genes, and generated sequence capture data from 99 RNA-seq libraries. We built a data
23 processing pipeline to quantify the RNA-CAP-I-seq data, and visualize differential gene
24 expression. Sequence capture in combination with quantitative RNA-seq enabled cost-
25 effective assessment of the expression profile of a specified subset of genes. Quantitative
26 sequence capture is not limited to RNA-seq or any specific organism and can potentially be
27 incorporated into automated platforms for high-throughput sequencing.

28

29 **KEYWORDS**

30 CAP-I, sequence capture, high-resolution expression profiling, quantitative RNA-seq, data
31 visualization, plant immunity, NLR, transcriptional regulation

32

---

[1] The Sainsbury Laboratory, University of East Anglia, Norwich Research Park, Norwich, NR4 7UH, United Kingdom
[2] Lead contact
[3] These authors contributed equally
[4] Present address: The University of North Carolina, Chapel Hill, North Carolina, 27599, United States of America
* Correspondence: pingtao.ding@tsl.ac.uk; jonathan.jones@tsl.ac.uk
Twitter: @sardineboy1 (PD), @BrunoNgou (BN), @jonathandgjones (JDGJ)

33 **RESULTS AND DISCUSSION**

34 In previous work, we investigated changes in *Arabidopsis thaliana* defense gene expression

35 in response to a bacterial effector after recognition via nucleotide-binding leucine-rich-repeat

36 intracellular immune receptors (NLRs). Specifically, we delivered the *Ralstonia*

37 *solanacearum* effector PopP2, and studied responses to its recognition by the RPS4/RRS1-

38 R intracellular immune receptor complex [9]. We defined a subset of early response genes

39 (ERGs) particularly responsive to NLR activation (Fig S1A, Table S1 and S2). Expression of

40 ERGs can be induced by both cell-surface receptors and NLRs, but more rapidly and

41 strongly induced when both classes of receptors are activated (Fig S1A). NLR-dependent

42 ERG upregulation was first observed at four hours post-infiltration (4 hpi) (Fig S1B, C). To

43 assess the roles of immune components during ERG activation, we measured ERG

44 transcripts in selected immune-deficient mutants compared to wild type (wt). Since these

45 studies involved multiple replicates, mutant backgrounds and treatments, we applied

46 complexity reduction via sequence capture to reduce sequencing costs.

47 We selected investigated 35 ERGs, and also 17 non-ERGs as controls, based on their

48 transcriptional regulation patterns (Fig S1A, Table S2) [9]. The ERGs include genes that are

49 important for conferring full resistance to various plant pathogens, and are involved in the

50 biosynthesis of phytohormones, salicylic acid (SA) and pipecolic acid (Pip), including *ICS1*,

51 *EDS5, PBS3, FMO1* and genes that encode the transcription factors (TFs) WRKY51 and

52 SARD1 [10–16, 17]. Non-ERG control genes include *UBQ10* and *ACT7*, as well as late

53 immune response genes [9], such as *PR1*, which is known to be activated by elevated SA

54 [18]. We included full-length gene loci as templates for the capture bait design, spanning

55 gene bodies (introns included) and putative promoters and terminators (Fig 1A). For

56 promoters and terminators, we either defined them based on the intragenic sequence region

57 between the coding sequence (CDS) of the target gene and the CDS of the immediate

58 neighboring genes (<4,500 base pairs, or bps), or used 4,500 bps upstream of the start

59 codon or downstream of the stop codon as their promoters or terminators, respectively (Fig

60 1A). This was to minimize the loss of any important sequence information: some genes

61 might need longer intragenic regions to be fully functional. All sequence templates were

62 designed using the gene coding strand (Fig 1A).

63 After computationally extracting sequences from all 52 gene loci, we used our bait design

64 pipeline to design a bait library (Fig 1A and Fig S2A). We synthesized a set of 20,000 120-

65 mer single-strand RNA probes (Fig 1A), which contains 2219 unique probes with 17-

66 nucleotide tiling and covering ~ 260 kb of the corresponding *Arabidopsis* genome regions

67 (Fig S1A). We named this library as 'Capture I' (CAP-I) for studies of plant innate immunity.

68 To test the efficiency of CAP-I for sequence capture, we performed one capture with libraries

69 generated from *Arabidopsis* genomic DNA for NGS. We found all gene loci have 100%

2

70   breadth of coverage (Fig 1B and Table S3), showing that CAP-I enables capture of targeted

71   sequences (Fig 1B). The pipeline generated one set of redundant baits in the region

72   between two adjacent genes (Fig S2B), which could be condensed to provide additional

73   capture capacity.

74   We then tested if CAP-I can be used in RNA-seq to assess quantitative changes in ERG

75   transcripts. We used *Arabidopsis thaliana* accession Col-0 as wt, and also investigated

76   seven selected mutants in Col-0 (Fig S2C). Resistance to *Ralstonia solanacearum* 1

77   (RRS1)-S and RRS1B are NLRs of bacterial effector AvrRps4, and they function together

78   with their paired NLRs Resistant to *Pseudomonas (P.) syringae* 4 (RPS4) and RPS4B,

79   respectively [19]; a *rrs1-3 rrs1b-1* mutant loses AvrRps4 responsiveness. EDS1 (the

80   included mutant is *eds1-2*) is required for immunity mediated by Toll/Interleukin-1

81   Receptor/Resistance (TIR)-NLRs like RRS1 and RPS4 [20]. *SID2* (the included mutant is

82   *sid2-2*) encodes the enzyme ICS1, which is required for the biosynthesis of defense-related

83   phytohormone, SA [10, 21]. SARD1 and its homolog Calmodulin-binding Protein 60-like g

84   (CBP60g) are master TFs required for transcriptional regulation of genes involved in

85   pathogen associated molecular pattern (PAMP)-triggered immunity (PTI), effector-triggered

86   immunity (ETI) and systemic acquired resistance (SAR) [13, 14, 22–24]. MYC2 and its

87   homologs MYC3 and MYC4 are basic helix-loop-helix TFs (the included mutant is *myc2*

88   *myc3 myc4*) required for jasmonic acid (JA)-mediated resistance against bacteria [25].

89   TOPLESS (TPL) and its homologs TPL-related 1 (TPR1) and TPR4 (the included mutant is

90   *tpl tpr1 tpr4*) are putative transcriptional co-repressors required for full resistance against the

91   bacterium *P. syringae* pv. *tomato* DC3000 (hereafter DC3000) and DC3000 expressing

92   AvrRps4 but not DC3000 expressing AvrRpt2, an effector recognized by RPS2, a non-TIR-

93   NLR [26]. Phytoalexin Deficient 4 (PAD4), Ethylene-insensitive protein 2 (EIN2), Delayed

94   Dehiscence 2 (DDE2, encoding an allene oxide synthase involved in jasmonic acid synthesis)

95   and SID2/ICS1 (the included mutant is *pad4-1 ein2-1 dde2-2 sid2-2*) are proteins that are

96   involved in different but interacting sectors in immune signaling networks [27].

97   Previously, we have defined the response induced by the bacterium *P. fluorescens* (Pf0-1

98   EtHAn strain) carrying a mutant effector PopP2^{C321A} (Pf0-1:PopP2^{C321A}) as 'PTI' mediated by

99   cell-surface Pathogen Recognition Receptors (PRRs) [9]. The Pf0-1 strain carrying wt

100   PopP2, recognized by RRS1-R/RPS4, triggers an additional ETI response that we designate

101   'PTI + ETI'. Here, we used Pf0-1:AvrRps4 or Pf0-1:AvrRpt2 to induce 'PTI + ETI'. The

102   responses induced by Pf0-1:AvrRps4 or AvrRpt2 are named as 'PTI plus TIR-NLR-mediated

103   ETI' (PTI + t-ETI) and 'PTI plus CC-NLR-mediated ETI' (PTI + c-ETI), respectively (Fig 3C).

104   In addition, Pf0-1 carrying the mutant effector AvrRps4$^{KRVY135-138AAAA}$ (Pf0-1:AvrRps4^{KRVYmut})

105   was included as 'PTI'. We also included leaves infiltrated with buffer only, as a mock

106   treatment, and no treatment on wt plants as an untreated control (Fig S2C). ERGs began to

107   show significant upregulation in their transcripts at 4 hpi of Pf0-1:PopP2 compared to Pf0-

108   1:PopP2[C321A] [9], so we collected our samples at 4 hpi for all treatments. For each

109   combination of genotype and treatment, we collected 3 biological replicates; 99 samples in

110   total (Fig S2C). We extracted RNAs from these samples and generated cDNA libraries. Each

111   library was barcoded with custom index primers. In addition, we added genomic DNA

112   libraries in the final multiplexed library as spike-in controls for sequence capture. We applied

113   one reaction of CAP-I baits to capture the multiplexed libraries before sequencing.

114   After demultiplexing, we retrieved single-end reads for each individual library. We

115   mapped the reads to CAP-I target gene loci and assessed the mapping efficiency. We

116   observed 100% breadth of coverage of full-length transcripts for all gene loci except for

117   *AT4G28410*, which encodes Root System Architecture 1 (RSA1). *RSA1* is specifically

118   expressed in *Arabidopsis* root tissue, and all our samples are leaf tissues, so *RSA1* served

119   as a good negative control for contamination introduced at any steps of library preparation

120   and sequencing. Since no reads from 99 cDNA libraries of RNA-CAP-I-seq mapped to the

121   *RSA1* locus while 100% breadth of coverage in *RSA1* locus occurred in the gDNA spike-in

122   controls (Fig S3A), it demonstrates our baits are specific and sensitive to any changes in the

123   quantity of targeted sequences. To test the reproducibility of each biological replicate, we

124   generated a sample correlation plot (Fig 2A). Results of three biological replicates from the

125   same combination of genotype and treatment group together based on their similarities, and

126   the average pair-wise correlation between them within groups is above 80% (Fig 2A). Thus,

127   the RNA-CAP-I-seq method is highly repeatable. To check how well our RNA-CAP-I-seq

128   captured differential gene expression, we visualized the mapped reads in a genome browser.

129   The overall expression pattern of *SARD1* gene in three biological replicates under all five

130   different treatments is similar (Fig 2B). More reads were mapped to *SARD1* in the samples

131   from "PTI", "PTI + t-ETI" and "PTI + c-ETI" than those in mock or untreated samples, which is

132   consistent with the previous observation of *SARD1* as one of the ERGs from the total RNA-

133   seq data [9]. Pathogen-induced SA accumulation is required for plant immunity, and one

134   major pathway of SA biosynthesis is *via* isochorismate (IC) [28]. The IC pathway involves

135   several enzymes that are required for the key catalytic steps, and encoded by *ICS1*, *EDS5*

136   and *PBS3* [29, 30]. They are all ERGs and directly regulated by TFs SARD1 and CBP60g [9,

137   23]. These three SA biosynthetic genes are usually transcriptionally co-regulated in the

138   activation of immunity and are also all highly induced in our 'PTI' and 'PTI + ETI' samples

139   (Fig 2C). Furthermore, 'PTI + ETI' induces stronger expression of these genes than 'PTI'

140   alone (Fig 2C), potentially through the regulation of SARD1 and CBP60g. In contrast, the

141   transcripts of the house-keeping genes, *UBQ10* and *ACT7* are stable regardless of the

142   treatments (Fig 2D).

143   Though we observed what we expected from the mapped reads, they required
144   normalization for statistical analysis of relative gene expression. For this, we have developed
145   an R package to normalize and visualize the data generated with sequence capture [31].
146   From the parameter of 'Goodness Of Fit', we found that not all selected control genes are
147   suitable for normalization as some of them are highly variable across 99 samples (Fig S3B).
148   After normalization, we obtained a balanced read distribution with low variation across all
149   samples (Table S4 and S5), enabling statistical analysis for differential gene expression. In
150   the clustering analysis, we retrieved three main clusters of genes based on their expression
151   patterns in all 32 different treatments compared to untreated Col-0 samples (Fig 3A, Table
152   S6). The majority of ERGs are in Cluster I and mostly are immunity related, while Cluster III
153   comprises predominantly control genes (Fig 3B, Table S7). Cluster II contains equal
154   numbers of ERGs and control genes (Fig 3A and 3B). From the same analysis, we also
155   identified three groups of conditions categorizing combinations of genotypes and treatments.
156   Regardless of the genotype, all mock treated samples are clustered in Group I with similar
157   expression patterns of CAP-I genes, indicating they serve as a good negative control for
158   other treatments. In Group III, overall expression of CAP-I genes had no discernable pattern
159   compared to that in Group I and II. In Group II, we were able to identify mutants that have
160   greater impacts on ERG expression pattern in response to treatments (Fig 3A). All Pf0-1-
161   treated samples in *sid2* mutant exhibit similar expression profiles, as do those in *sard1*
162   *cbp60g* double mutant. These indicate that ICS1 or SARD1/CBP60g are required for the
163   activation of both 'PTI' and 'PTI + ETI'. Consistent with EDS1 being required for AvrRps4-
164   but not AvrRpt2-induced ETI, our results also show that ERGs in *eds1* are induced less by
165   Pf0-1:AvrRps4 and Pf0-1:AvrRps4[KRVYmut] (eds1_a4 and eds1_kv) in comparison to those
166   induced by Pf0-1:AvrRpt2 (eds1_a2) (Fig 3A). We also observed that ERGs are induced less
167   in a *pad4 ein2 dde2 sid2* quadruple mutant (*peds*) than in wt by 'PTI', which is consistent
168   with previous reports [27, 32]. However, we did not see a strong ERG difference between
169   *peds* and wt in response to 'PTI + ETI' (Fig 3A).

170   t-ETI and c-ETI confer resistance *via* different types of NLRs and signaling components [8,
171   20]. However, there is no previously reported side-by-side comparison of TIR-NLR- and CC-
172   NLR-induced genes upon NLR activation. Here, we compared the induction patterns of
173   ERGs in wt treated with 'PTI + t-ETI' and 'PTI + c-ETI', and they significantly resemble each
174   other for all CAP-I genes (R^2=0.81) (Fig 3C). As the 32 conditions are combinations of both
175   genotypes and treatments, we checked the correlation of gene expression patterns with
176   either genotypes or treatments separately (Fig 4A). Gene expression patterns from the
177   treatments of 'PTI + t-ETI' and 'PTI + c-ETI' within the same genotype tend to group together,
178   rather than with 'PTI' (Fig 4A), which further proves that gene expression patterns induced
179   by TIR-NLRs and CC-NLRs at early immune activation stages are similar.

180   We examined differential gene expression between each individual mutant and wt. As

181   expected, in both *eds1* and *rrs1 rrs1b* mutants, gene expression patterns are similar

182   between the two treatments of Pf0-1:AvrRps4 and Pf0-1:AvrRps4^{KRVYmut}, because both

183   EDS1 and RRS1/RRS1B are required for AvrRps4-induced ETI. Loss-of-function of the

184   AvrRps4 receptors (*rrs1 rrs1b*) or the downstream signaling component EDS1 (*eds1*)

185   resemble the loss-of-recognition of AvrRps4 due to the mutation of AvrRps4 (Pf0-

186   1:AvrRps4^{KRVYmut}) in wt plants (Fig 4B and 4C). On the other hand, EDS1 and RRS1/RRS1B

187   are not required for AvrRpt2 recognition, so Pf0-1:AvrRpt2 can still induce both PTI and ETI

188   in *eds1* and *rrs1 rrs1b* mutants (Fig 4B and 4C).

189   The TFs SARD1 and CBP60g bind to the promoters of defense genes to regulate their

190   expression [13, 23]. We observed that most ERGs that are downregulated in *sard1 cbp60g*

191   mutants are also identified as targets of SARD1 from chromatin immunoprecipitation

192   followed by sequencing (ChIP-seq) of SARD1 (Fig 4D) [23].

193   The *sid2* mutant is known to have no expression of the *ICS1* gene and compromised SA

194   accumulation induced by pathogens, so we expected to see that SA-induced genes were

195   also downregulated. We observed that genes induced by SA and upregulated during SAR,

196   specifically *PR1 and Acireductone Dioxygenase 3* (*ARD3*) were both downregulated in *sid2*

197   (Fig S4A). *SARD1* is also downregulated in *sid2*, indicating that SARD1-dependent

198   regulation of *ICS1* and SA biosynthesis can in turn positively regulate *SARD1* gene

199   expression. TF WRKY51 and its homolog WRKY50 positively regulate SA signaling and

200   negatively regulate JA signaling [17]. In *wrky50 wrky51* loss-of-function mutants, *Plant*

201   *Defensin 1.2A* (*PDF1.2A*) is downregulated in response to JA [17]. Here, we found in a *sid2*

202   mutant, *WRKY51* is downregulated, while *PDF1.2A* is upregulated (Fig S4A), which is

203   consistent with the negative expression association between *WRKY51* and *PDF1.2A*. In

204   addition, we found *Cytochrome P450 Monooxygenase 71A13* (*CYP71A13*) is downregulated

205   in *sid2* upon activation of innate immunity, indicating that SA might play positive regulatory

206   roles in camalexin synthesis [33].

207   The expression of JA response genes *Tyrosine Aminotransferase 3* (*TAT3*) and

208   *Lipoxygenase 2* (*LOX2*) but not *PDF1.2A* is positively regulated by MYC2 and its

209   homologues MYC3 and MYC4 [25, 34]. In our RNA-CAP-I-seq data, we found *MYC2, TAT3*

210   and *LOX2* are downregulated in *myc2 myc3 myc4* triple mutants, whereas *PDF1.2A* is

211   upregulated in the triple mutant in response to activation of innate immunity (Fig S4B).

212   TOPLESS mutants *tpr1 tpl tpr4* show enhanced susceptibility to bacteria DC3000 and

213   DC3000 carrying AvrRps4 [26]. However, this cannot be simply explained by the expression

214   pattern of ERGs, as we found no clear reduction of ERGs in *tpr1 tpl tpr4* mutants (Fig S5C).

215   Previously TOPLESS proteins were reported as transcriptional co-repressors, but there is

216   only slight evidence in our data of TOPLESS repressor activity towards a few specific genes.

217 Here, we found some defense-related ERGs are downregulated, while others are
218 upregulated, in response to both 'PTI + t-ETI' and 'PTI + c-ETI' compared to 'PTI', which
219 indicates that TOPLESS proteins may play dual functions or indirect roles in regulating
220 ERGs. As there is no ChIP-seq data of TOPLESS proteins or related histone modification
221 marks available, their functions remain unclear. Our data, together with previous reports,
222 nevertheless indicate a complex contribution of TOPLESS proteins in regulating genes
223 induced during plant immunity (Fig S4C) [26].
224   The *peds* mutant carries mutations in genes from four major immune sectors: PAD4
225 (*pad4*), ethylene (*ein2*), JA (*dde2*) and SA (*sid2*) [27]. We observed that PAD4, SA and JA
226 response genes are downregulated in *peds*, including *PAD4, ICS1*, *EDS5*, *WRKY51*,
227 *CYP71A13*, *MYC2*, *TAT3* and *LOX2* (Fig S4D). It has been reported that the *PEDS*-
228 represented phytohormone network is required for achieving higher amplitude of
229 transcriptional reprogramming during early CC-NLR-activated ETI in addition to PTI than
230 during PTI alone [35]. However in that report [35], the authors used DC3000 instead of Pf0-1
231 in our case, which can not only trigger 'PTI + ETI' but the background effectors in DC3000
232 can also trigger effector-triggered susceptibility ('ETS'), so our results using Pf0-1 are
233 'cleaner'. We showed a greater expression difference of ERGs activated by 'PTI' and by 'PTI
234 + ETI' in *peds* mutant compared to wt (Fig S4D). Like AvrRpt2, AvrRpm1 is also recognized
235 by a CC-NLR, Resistance to *P. syringae* pv *maculicola* 1 (RPM1) and activates ETI [8, 36].
236 Unlike AvrRpt2-induced ETI, AvrRpm1-induced ETI does not require *PEDS*-represented
237 phytohormone network to achieve a high-amplitude transcriptional reprogram within the early
238 time window of ETI activation [35]. Data from the same report indicate that *RPS2*, but not
239 *RPM1*, gene expression is highly reduced in *peds* when ETI was activated [35]. From this we
240 hypothesize that *RPS2* gene expression might be regulated through these four sectors,
241 explaining why all AvrRpt2-induced ERGs are delayed in contrast to AvrRpm1-induced ETI.
242   Here, using a limited subset of genes (CAP-I), we could distinguish gene expression
243 profiles during 'PTI', 'PTI + c-ETI', 'PTI + t-ETI' in various mutants, particularly the immune
244 gene regulatory components EDS1, ICS1 and SARD1/CBP60g. Inclusion of additional
245 innate immunity genes in the bait library should enable us to distinguish mutants with
246 enhanced resolution. In addition, as all steps for CAP-I are easy to follow and reproducible,
247 CAP-seq can be further implemented in an automated platform for more high-throughput
248 applications.
249 Single cell RNA-seq (scRNA-seq) for signature genes is available for some plant tissues [37,
250 38], and could be combined with capture-seq. A set of 100 marker genes has been defined
251 for *Arabidopsis* that can be used to predict the total transcriptome for each species [39];
252 these could be incorporated into future capture-seq bait library design. Capture-seq is also
253 capable of comparing the changes in the abundance of any DNA sequences, so it is not

254   limited to cDNA libraries, but can be used in other types of DNA libraries, such as ChIP-seq

255   and Assay for Transposase-Accessible Chromatin using sequencing (ATAC-seq) [40, 41].

256   Finally, capture-seq could also be used to investigate expression of specific pathogen genes

257   during host colonization (Pathogen Enrichment Sequencing: PenSeq) [6, 7]. In summary,

258   sequence capture provides an extremely versatile and cost-effective method to investigate

259   changes in expression of any designated gene set.

260

261   **FIGURE LEGENDS**

262   **Figure 1 CAP-I Bait Design and Validation**

263   (A) Visualization of bait design on one of CAP-I gene loci, *SARD1*. Using GFF file, here we

264   present the genome organization of one CAP-I gene locus, *SARD1*. Top row shows the

265   annotated exons and introns and intragenic regions of CAP-I gene locus and neighboring

266   gene loci. Second row show the direction of the coding strand, here *SARD1* coding is on the

267   reverse strand. The third row shows the orientation and the region that covers *SARD1* loci

268   and putative promoter and terminator. The fourth strand shows the final non-redundant baits

269   we designed and how they are mapped to the CAP-I target gene locus. The final baits are

270   120 nucleotides (nt) in length with 17 nt overlap for tilling.

271   (B) Trial run of CAP-I-seq reads from genomic DNAs mapped to *SARD1* locus and

272   visualized in a genome browser. Illumina sequencing reads of genomic DNA (gDNA) with

273   four biological replicates in one CAP-I capture shows 100% coverage on all CAP-I gene loci

274   including *SARD1*.

275   See also Figure S1, Table S1 to S3.

276   **Figure 2 Reproducibility Test of RNA-CAP-I-seq**

277   (A) Correlation analysis of mapped reads from all individual libraries from RNA-CAP-I-seq.

278   All individual libraries including cDNA libraries and spiked-in gDNA libraries from the same

279   CAP-I-seq are pair-wisely compared. 1 indicates 100% positive correlation based on the

280   distribution of reads, while -1 indicates 100% negative correlation.

281   (B) to (D) Mapped reads before normalization are visualized in several CAP-I gene loci in a

282   genome browser.

283   (B) Visualization of reads mapped to *SARD1* locus from wt samples. All three biological

284   replicates (r1-r3) of wt plants under five different treatments are visualized in IGV genome

285   browser at *SARD1* locus. Black indicates untreated (un); orange indicates samples collected

286   at 4 hour post infiltration (hpi) of mock (10 mM $MgCl_2$) treatment (mk); sky blue indicates

287   samples collected at 4 hpi of Pf0-1:AvrRps4[KRVYmut] (kv); bluish green indicates samples

288   collected at 4 hpi of Pf0-1:AvrRps4 (a4); vermilion indicates samples collected at 4 hpi of

289   Pf0-1:AvrRpt2 (a2).

290   See also Figure S2.

291 **Figure 3 Quantification of RNA-CAP-I-seq**

292 (A) Cluster analysis of normalized read counts from each combination of conditions in

293 comparison to untreated wt Col-0 samples (wt_un). Each combination of conditions

294 represents all combinations of each genotype (wt, eds1, r1ab, sid2, gh, myc234, tplr14, peds)

295 with each treatment (mk, kv, a4, a2). CAP-I genes form three major clusters based on their

296 expression patterns cross all conditions. All conditions form three major groups based on

297 their overall differential gene expression of CAP-I genes. ERGs from CAP-I are in orange

298 and control genes are in sky blue. Heatmap is based on mean z-scores of three biological

299 replicates. Redder color indicates a higher value of z-score, while bluer means a less value

300 of z-score.

301 (B) Top hits of gene ontology (GO) terms based on their p-values for CAP-I genes in each

302 cluster from (A). BP stands for biological process, KEGG is based on the database from

303 Kyoto Encyclopedia of Genes and Genomes. WP refers to WikiPathways database.

304 (C) Comparison of differential gene expression patterns of all CAP-I genes activated by ETI

305 between RRS1/RPS4 and RPS2 in addition to PTI.

306 See also Figure S3, Table S4 to S7.

307 **Figure 4 Correlation studies of RNA-CAP-I-seq from different genotypes and**

308 **treatments.**

309 (A) Correlation analysis with mapped and normalized reads from 32 different combinations

310 of both genotypes and treatments. For treatments, we use color-filled circles to indicate,

311 Black circles stand for mock treatment. Sky blue circles are for Pf0-1:AvrRps4^{KRVYmut} (kv).

312 Vermilion circles are for Pf0-1:AvrRps4 or AvrRpt2. For genotypes, we use color-filled

313 squares to indicate, Black squares are for wt Col-0. Yellow squares are for *rrs1-1 rrs1b-1*

314 double mutants. Orange squares are for *eds1-2* (Col-0) mutant. Reddish purple squares are

315 for *sid2-2* mutant, Vermilion squares are for *pad4-1 ein2-1 dde2-2 sid2-2* quadruple mutants.

316 Sky blue stands for *sard1-1*. Bluish green stands for *myc2/3/4, Blue* is for TOPLESS mutants

317 *tprl tpr1 trpr4*.

318 (B) to (D) Differential gene expression are visualized with heat maps. (B) Heatmap of

319 differential expression of CAP-I genes in *rrs1 rrs1b* double mutants compared to wt. (C)

320 Heatmap of CAP-I genes in *eds1* mutant compared to wt. (D) Heatmap of CAP-I genes in

321 sard1 cbp60g mutants compared to wt.

322 See also figure S4.

323

324 **Supplemental Information**

325 **Figure S1 Time-series expression of CAP-I genes under different conditions of**

326 **immune activation.**

9

327 (A) Heatmap of z-scores for CAP-I genes based on their time-series read counts in the total

328 RNA-seq. ERGs are in orange and selected control genes are in sky blue. Based on the

329 clustering analysis, all ERGs are grouped in one clade, while control genes are grouped in

330 another clade except for two genes.

331 (B) and (C) Time-series visualization of the mean value of z-scores for ERGs and control

332 genes. (B) ERGs show overall upregulations under PTI + ETI treatment compared to PTI

333 alone. (C) Control genes didn't show such pattern as (B), but overall are stable in their

334 expression level during the course of both immune activations.

335 **Figure S2 CAP-I bait design and RNA-CAP-I-seq experimental design.**

336 (A) Pipeline for CAP-I bait library design. All script information and original files are available

337 in our Github via: https://github.com/slt666666/Ding_etal_2019_CAP_I

338 (B) Visualization of one duplicated region for CAP-I bait design. AT1G53620 and

339 AT1G53625 are two ERG loci next to each other in the genome. The orange overlapped

340 squares with filled orange color are baits for AT1G53620, and those squares without filling

341 color are baits for AT1G53625. Because they are neighboring genes on the genome, so we

342 have got two sets of baits for the overlapped region between these two genes.

343 (C) Experimental design for RNA-CAP-I-seq. Plants from eight different genotypes including

344 wild-type Col-0 accession are treated with four different conditions, which generates 32

345 different combinations. Untreated Col-0 wt plants are included as an additional control

346 condition, so there are 33 different conditions. All combinations have 3 biological replicates

347 for later-on statistics, so in total, we have 99 individual libraries for CAP-I-seq.

348 **Figure S3 Overall quality assessment of RNA-CAP-I-seq data.**

349 (A) RNA-CAP-I-seq reads mapped to control gene *RSA1*. There are nearly no reads

350 mapped to *RSA1* in all 99 cDNA sequencing results, while there are 100% coverage of

351 reads at *RSA1* locus from gDNA. Here we show mapped reads from five different cDNA

352 libraries as examples with the gDNA mapped reads as control.

353 (B) List of control genes for normalizing overlapped with CAP-I control gene set. Density plot

354 with Goodness Of Fit (GOF) shows 9 selected control genes that are much less variable.

355 (C) Comparison of RNA-CAP-I-seq data with previous published RNA-seq data with similar

356 conditions. Ws-0_0hr in previous publication is treated equivalently as untreated wt (wt_un)

357 in this study. Ws-0_4hr-PopP2 stands for 4 hpi of Pf0-1:PopP2 in wt Ws-0, which activates

358 ETI via RRS1-R/RPS4 in addition to PTI; this is treated equivalently as wt_a4, as this

359 activates ETI via RRS1-S/RPS4 in addition to PTI. Similarly, C321A stands for mutant

360 PopP2 and kv stands for mutant AvrRps4, in which case, both only activate PTI. The

361 comparison is based on the normalized read counts from both datasets.

362 (D) Pair-wise comparisons of CAP-I gene expressions from AvrRpt4-, AvrRpt2- and PopP2-

363 induced wt samples.

10

364   Figure S4 Heatmaps of differential gene expression of mutants individually compared to wt.

365   (A) to (E) are heatmaps based on z-scores, they are for *sid2*, *myc2/myc3/myc4*, *tpr1 tpl tpr4*

366   and *pad4 ein2 dde2 sid2* individually compared with wt.

367

368   **Table S1 Information_of_ERGs_and_Control_Genes_in_CAP-I**

369   **Table S2 Time-series_Expression_of_CAP-I_genes**

370   **Table S3 Coverage_Information_of_CAP-I_gDNA_seq_Trial**

371   **Table S4 Read_Counts_of_RNA-CAP-I-seq_before_Normalisation**

372   **Table S5 Read_Counts_of_RNA-CAP-I-seq_post_Normalisation**

373   **Table S6 Log_Matrix_of_RNA-CAP-I-seq_Normalised_to_wt_un**

374   **Table S7**

375   **Geno_Ontology_Information_for_Clusters_in_Differential_Gene_Expression_Heatmap**

376   **Table S8 Barcode_Information_for_RNA-Cap-I_Seq**

377

378   ***Plant material and growth condition***

379   Mutants of *rrs1-3 rrs1b-1*, *eds1-2*, *sid2-2*, *sard1-1 cbp60g-1*, *myc2 myc3 myc4*, *tpr1 tpl tpr4*

380   and *pad4-1 ein2-1 dde2-2 sid2-2* that were used in this study have been previously

381   described [13, 19, 25–27, 42, 43]. Seeds were sown on compost and plants were grown at

382   21°C with 10 hours under light and 14 hours in dark, and at 70% humidity.

383

384   **METHOD DETAILS**

385   ***Bacterial infiltration assay and sample collection***

386   All Pf0-1 strains with different effectors were streaked from their glycerol stock in -70°C

387   freezer on petri dish plates with King's B medium containing antibiotics for positive selection.

388   Pf0-1:AvrRps4 and Pf0-1:AvrRps4^{KRVYmut} positive colonies were selected with 5 μg/ml

389   tetracycline, 10 μg/ml chloramphenicol and 20 μg/ml gentamycin. Pf0-1:AvrRpt2 were

390   selected with 5 μg/ml tetracycline, 10 μg/ml chloramphenicol and 10 μg/ml kanamycin.

391   Plates were growing in 28°C thermo incubator overnight. Fresh bacteria were streaked off

392   from the plate surface with 1ml clean pipette tips and resuspended in freshly prepared sterile

393   10 mM MgCl$_2$, and spun with 5, 000 rpm for 3 minutes at room temperature. Discarded the

394   supernatant and resuspended the pellet with 10 mM MgCl$_2$. The concentration of bacteria

395   was measured and indicated with the optical density at a wavelength of 600 nm (OD$_{600}$).

396   Final concentration of OD$_{600}$=0.2 were used for infiltration with 1 ml needleless syringes. 2

397   fully expanded leaves from a 5-week-old plant were infiltrated with one of the bacterial

398   strains or just 10 mM MgCl$_2$ resuspending buffer as mock. Six leaves from three plants were

399   collected at 4 hours post infiltration (hpi) for each genotype under one certain treatment.

400 Leaves are snap frozen in liquid nitrogen for following up RNA extraction. Three batches of
401 plants were grown under the same condition but on different dates, and samples collected
402 from these three batches are used as three biological replicates.

403

404 **RNA extraction**

405 All samples were kept in -70°C freezer from liquid nitrogen if the RNAs were not extracted
406 immediately after sample collection. Total RNAs were extracted with Quick-RNA™ Plant
407 Miniprep Kit (Catalog No. R2024, Zymo Research) following the protocol provided by Zymo
408 Research. The quantities of RNAs were measured by Nanodrop and the qualities of RNAs
409 were assessed with the RNA 6000 Nano Kit (Catalog No. 5067-1511) on an Agilent 2100
410 Bioanalyzer System. mRNAs were purified with 2 times of enrichment using Dynabeads™
411 Oligo (dT)25 (Catalog No. 61002, Invitrogen™) from the total RNAs. The qualities and
412 quantities of mRNAs were assessed with the RNA 6000 Pico Kit (Catalog No. 5067-1513,
413 Agilent) on an Agilent 2100 Bioanalyzer System.

414

415 **cDNA library construction for RNA-CAP-I-seq**

416 mRNAs were submitted for first strand synthesis with Random Decamers (50 µM) (Catalog
417 No. AM5722G, Invitrogen™) and SuperScript™ IV Reverse Transcriptase kit (Catalog No.
418 18090200, Invitrogen™). The second strand cDNA synthesis was carried out as previously
419 described [44, 45]. Concentration of double strand cDNAs were quantified with the HS
420 dsDNA Assay kit (Catalog No. Q32851, Invitrogen™) on a Qubit Fluorometer. Illumina
421 sequencing-compatible cDNA libraries were constructed using tagmentation [46]. All libraries
422 were barcoded with in-house custom designed primers (Table S8) and assessed with the
423 High Sensitivity DNA Kit (Catalog No. 5067-4626, Aligent) on an Agilent 2100 Bioanalyzer
424 System.

425

426 **CAP-I bait design and RNA-CAP-I sequence capture**

427 For enrichment of selected ERGs and controls, 2219 synthetic 120-nt biotinylated RNA
428 probes with 17 bp tiling were designed and synthesized, complementary to 52 gene regions
429 (including promoter, coding, intron and terminators) totaling 261,616 bp from the reference
430 genome of *Arabidopsis thaliana* Col-0 [47] (MYbaits; MYcroarray now is Arbor Biosciences,
431 MI, USA; https://arborbiosci.com/). Repetitive regions of total 18800 bp within the targeted
432 sequences were masked using RepeatMasker (Smit AFA, Hubley R & Green P.
433 RepeatMasker Open-4.0. 2013-2015), and two highly represented baits with >10
434 MEGABLAST hits to the TAIR10 reference genome were removed [48]. All detailed
435 information can also be found in our GitHub (Link). In preparation for sequencing, barcoded
436 libraries were sized on the Agilent 2100 Bioanalyzer and then quantified using the Qubit

12

437 Fluorometer and real-time quantitative PCR (Catalog no. KK4824, Kapa Biosystems).
438 Individual samples were pooled equimolarly. After multiplexing, the RNA-CAP-I library was
439 carried out for sequence capture with CAP-I baits following the protocol provided with
440 blockers specifically for indices with 9 nucleotides. (https://arborbiosci.com/wp-
441 content/uploads/2017/10/MYbaits-manual-v3.pdf)

442

443 **RNA-CAP-I-seq on a NextSeq 500 sequencer**
444 The multiplexed libraries were used as input following the NextSeq 500 instrument sample
445 preparation protocol (Catalog no. 15048776, Illumina). With a recommended 1.8-pM library
446 concentration resulted in clustering density in our instrument (276,000 clusters/mm2).
447 Samples were sequenced on a single flow cell of the NextSeq 500/550 High Output kit (75
448 cycles), using a 74-cycle (single-end) configuration. The sequencing run in the NextSeq 500
449 produced over 600 million single-end reads with a Q30 ≥ 92.5%.

450

451 **Demultiplexing raw data from the NextSeq 500**
452 Raw sequence data obtained from Illumina NextSeq500 sequencing platform are per-cycle
453 base call (BCL) format. As many analysis application tools require per-read FASTQ format
454 files as an input, we need to transform bcl file to fastq. A conversion software by Illumina
455 called bcl2fastq version 2.20.0 (http://emea.support.illumina.com/downloads/bcl2fastq-
456 conversion-software-v2-20.html), was used to demultiplex samples and convert the BCL
457 format to FASTQ format. A sample sheet was prepared following the user guide
458 (https://support.illumina.com/content/dam/illumina-
459 support/documents/documentation/software_documentation/bcl2fastq/bcl2fastq2-v2-20-
460 software-guide-15051736-03.pdf). The sample sheet contains sample identifier and a
461 barcode or a barcode pair (nucleotide bases) and is provided to bcl2fastq for correct
462 demultiplexing of the sample sequence reads. More detail about the command line usage of
463 bcl2fastq tool can be obtained in the user guide. All raw reads post demultiplexing will be
464 open access through the European Nucleotide Archive (ENA) under the accession number
465 of PRJEB34520.

466

467 **Mapping reads to genome data, transcript annotation, and profiling of gene**
468 **expression**
469 The single-end reads for cDNA libraries were mapped to the *Arabidopsis thaliana* Col-0
470 reference genome (TAIR10) using TopHat v.2.1.1 [49]. Reads from the spike-in genomic
471 DNA were aligned to the reference genome using Bowtie2 v2.2.9[50]. The resulting BAM
472 files were sorted with SAMtools before downstream analysis [51]. With sorted BAM files, all
473 downstream analysis following the pipeline of 'atacR' [31]. All the data that we were not able

474   to include in the supplemental materials are available in Github

475   (https://github.com/slt666666/Ding_etal_2019_CAP_I). All scripts and files we generated for

476   this study are available in our Github (https://github.com/slt666666/Ding_etal_2019_CAP_I).

477

489

490   **Author Contributions**

491   PD and JDGJ conceptualized the study. Experiments and methods of RNA-CAP-I-seq are

492   designed by PD, and are carried out by PD and BN. CAP-I baits and barcodes are designed

493   by PD and OJF. Data analysis are performed by PD, DM, RKS and TS. The original draft

494   was written by PD, and reviewed and edited by JDGJ, BN, OJF, TS, RKS and DM.

495

496   **Declaration of Interests**

497   The authors declare no competing interests.

498

499   **Reference**

500   1.   Jupe, F., Witek, K., Verweij, W., Sliwka, J., Pritchard, L., Etherington, G. J., Maclean, D.,
501   Cock, P. J., Leggett, R. M., Bryan, G. J., et al. (2013). Resistance gene enrichment
502   sequencing (RenSeq) enables reannotation of the NB-LRR gene family from sequenced
503   plant genomes and rapid mapping of resistance loci in segregating populations. Plant J.
504   *76*, 530–544.

505   2.   Mercer, T. R., Clark, M. B., Crawford, J., Brunck, M. E., Gerhardt, D. J., Taft, R. J.,
506   Nielsen, L. K., Dinger, M. E., and Mattick, J. S. (2014). Targeted sequencing for gene
507   discovery and quantification using RNA CaptureSeq. Nat. Protoc. *9*, 989–1009.

508   3.   Witek, K., Jupe, F., Witek, A. I., Baker, D., Clark, M. D., and Jones, J. D. G. (2016).
509   Accelerated cloning of a potato late blight-resistance gene using RenSeq and SMRT
510   sequencing. Nat. Biotechnol. *34*, 656–660.

511   4.   Witek, K., Witek, K., Jupe, F., Witek, A. I., Baker, D., Clark, M. D., and Jones, J. D.

512   (2016). SMRT RenSeq protocol. Protoc exch.

5.   Van de Weyer, A.-L., Monteiro, F., Furzer, O. J., Nishimura, M. T., Cevik, V., Witek, K., Jones, J. D. G., Dangl, J. L., Weigel, D., and Bemm, F. (2019). A Species-Wide Inventory of NLR Genes and Alleles in Arabidopsis thaliana. Cell *178*, 1260–1272.e14.

6.   Jouet, A., Saunders, D. G. O., McMullan, M., Ward, B., Furzer, O., Jupe, F., Cevik, V., Hein, I., Thilliez, G. J. A., Holub, E., et al. (2019). Albugo candida race diversity, ploidy and host-associated microbes revealed using DNA sequence capture on diseased plants in the field. New Phytol. *221*, 1529–1543.

7.   Thilliez, G. J. A., Armstrong, M. R., Lim, T.-Y., Baker, K., Jouet, A., Ward, B., van Oosterhout, C., Jones, J. D. G., Huitema, E., Birch, P. R. J., et al. (2019). Pathogen enrichment sequencing (PenSeq) enables population genomic studies in oomycetes. New Phytol. *221*, 1634–1648.

8.   Jones, J. D. G., and Dangl, J. L. (2006). The plant immune system. Nature *444*, 323–329.

9.   Sohn, K. H., Segonzac, C., Rallapalli, G., Sarris, P. F., Woo, J. Y., Williams, S. J., Newman, T. E., Paek, K. H., Kobe, B., and Jones, J. D. G. (2014). The nuclear immune receptor RPS4 is required for RRS1SLH1-dependent constitutive defense activation in Arabidopsis thaliana. PLoS Genet. *10*, e1004655.

10. Wildermuth, M. C., Dewdney, J., Wu, G., and Ausubel, F. M. (2001). Isochorismate synthase is required to synthesize salicylic acid for plant defence. Nature *414*, 562–565.

11. Nawrath, C., Heck, S., Parinthawong, N., and Métraux, J.-P. (2002). EDS5, an essential component of salicylic acid-dependent signaling for disease resistance in Arabidopsis, is a member of the MATE transporter family. Plant Cell *14*, 275–286.

12. Nobuta, K., Okrent, R. A., Stoutemyer, M., Rodibaugh, N., Kempema, L., Wildermuth, M. C., and Innes, R. W. (2007). The GH3 acyl adenylase family member PBS3 regulates salicylic acid-dependent defense responses in Arabidopsis. Plant Physiol. *144*, 1144–1156.

13. Zhang, Y., Xu, S., Ding, P., Wang, D., Cheng, Y. T., He, J., Gao, M., Xu, F., Li, Y., Zhu, Z., et al. (2010). Control of salicylic acid synthesis and systemic acquired resistance by two members of a plant-specific family of transcription factors. Proc. Natl. Acad. Sci. USA *107*, 18220–18225.

14. Wang, L., Tsuda, K., Truman, W., Sato, M., Nguyen, L. V., Katagiri, F., and Glazebrook, J. (2011). CBP60g and SARD1 play partially redundant critical roles in salicylic acid signaling. Plant J. *67*, 1029–1041.

15. Hartmann, M., Zeier, T., Bernsdorff, F., Reichel-Deland, V., Kim, D., Hohmann, M., Scholten, N., Schuck, S., Bräutigam, A., Hölzel, T., et al. (2018). Flavin Monooxygenase-Generated N-Hydroxypipecolic Acid Is a Critical Element of Plant Systemic Immunity. Cell *173*, 456–469.e16.

16. Chen, Y. C., Holmes, E., Rajniak, J., Kim, J.-G., Tang, S., Fischer, C., Mudgett, M. B., and Sattely, E. (2018). N-hydroxy-pipecolic acid is a mobile signal that induces systemic disease resistance in Arabidopsis. BioRxiv.

17. Gao, Q.-M., Venugopal, S., Navarre, D., and Kachroo, A. (2011). Low oleic acid-derived repression of jasmonic acid-inducible defense responses requires the WRKY50 and

555   WRKY51 proteins. Plant Physiol. *155*, 464–476.

556 18. Cao, H., Glazebrook, J., Clarke, J. D., Volko, S., and Dong, X. (1997). The Arabidopsis
557   NPR1 gene that controls systemic acquired resistance encodes a novel protein
558   containing ankyrin repeats. Cell *88*, 57–63.

559 19. Saucet, S. B., Ma, Y., Sarris, P. F., Furzer, O. J., Sohn, K. H., and Jones, J. D. G.
560   (2015). Two linked pairs of Arabidopsis TNL resistance genes independently confer
561   recognition of bacterial effector AvrRps4. Nat. Commun. *6*, 6338.

562 20. Aarts, N., Metz, M., Holub, E., Staskawicz, B. J., Daniels, M. J., and Parker, J. E. (1998).
563   Different requirements for EDS1 and NDR1 by disease resistance genes define at least
564   two R gene-mediated signaling pathways in Arabidopsis. Proc. Natl. Acad. Sci. USA *95*,
565   10306–10311.

566 21. Dewdney, J., Reuber, T. L., Wildermuth, M. C., Devoto, A., Cui, J., Stutius, L. M.,
567   Drummond, E. P., and Ausubel, F. M. (2000). Three unique mutants of Arabidopsis
568   identify eds loci required for limiting growth of a biotrophic fungal pathogen. Plant J. *24*,
569   205–218.

570 22. Wang, L., Tsuda, K., Sato, M., Cohen, J. D., Katagiri, F., and Glazebrook, J. (2009).
571   Arabidopsis CaM binding protein CBP60g contributes to MAMP-induced SA
572   accumulation and is involved in disease resistance against Pseudomonas syringae.
573   PLoS Pathog. *5*, e1000301.

574 23. Sun, T., Zhang, Y., Li, Y., Zhang, Q., Ding, Y., and Zhang, Y. (2015). ChIP-seq reveals
575   broad roles of SARD1 and CBP60g in regulating plant immunity. Nat. Commun. *6*,
576   10159.

577 24. Ding, P., and Redkar, A. (2018). Pathogens suppress host transcription factors for
578   rampant proliferation. Trends Plant Sci. *23*, 950–953.

579 25. Fernández-Calvo, P., Chini, A., Fernández-Barbero, G., Chico, J. M., Gimenez-Ibanez,
580   S., Geerinck, J., Eeckhout, D., Schweizer, F., Godoy, M., Franco-Zorrilla, J. M., et al.
581   (2011). The *Arabidopsis* bHLH transcription factors MYC3 and MYC4 are targets of JAZ
582   repressors and act additively with MYC2 in the activation of jasmonate responses. Plant
583   Cell *23*, 701–715.

584 26. Zhu, Z., Xu, F., Zhang, Y., Cheng, Y. T., Wiermer, M., Li, X., and Zhang, Y. (2010).
585   Arabidopsis resistance protein SNC1 activates immune responses through association
586   with a transcriptional corepressor. Proc. Natl. Acad. Sci. USA *107*, 13960–13965.

587 27. Tsuda, K., Sato, M., Stoddard, T., Glazebrook, J., and Katagiri, F. (2009). Network
588   properties of robust immunity in plants. PLoS Genet. *5*, e1000772.

589 28. Dempsey, D. A., Vlot, A. C., Wildermuth, M. C., and Klessig, D. F. (2011). Salicylic Acid
590   biosynthesis and metabolism. Arabidopsis Book *9*, e0156.

591 29. Rekhter, D., Lüdke, D., Ding, Y., Feussner, K., Zienkiewicz, K., Lipka, V., Wiermer, M.,
592   Zhang, Y., and Feussner, I. (2019). Isochorismate-derived biosynthesis of the plant
593   stress hormone salicylic acid. Science *365*, 498–502.

594 30. Torrens-Spence, M. P., Bobokalonova, A., Carballo, V., Glinkerman, C. M., Pluskal, T.,
595   Shen, A., and Weng, J.-K. (2019). PBS3 and EPS1 complete salicylic acid biosynthesis
596   from isochorismate in Arabidopsis. BioRxiv.

597 31. Shrestha, R. K., Ding, P., Jones, J. D. G., and MacLean, D. (2018). A workflow for
598   simplified analysis of ATAC-cap-seq data in R. Gigascience *7*.

599 32. Hillmer, R. A., Tsuda, K., Rallapalli, G., Asai, S., Truman, W., Papke, M. D., Sakakibara,
600   H., Jones, J. D. G., Myers, C. L., and Katagiri, F. (2017). The highly buffered
601   Arabidopsis immune signaling network conceals the functions of its components. PLoS
602   Genet. *13*, e1006639.

603 33. Nafisi, M., Goregaoker, S., Botanga, C. J., Glawischnig, E., Olsen, C. E., Halkier, B. A.,
604   and Glazebrook, J. (2007). Arabidopsis cytochrome P450 monooxygenase 71A13
605   catalyzes the conversion of indole-3-acetaldoxime in camalexin synthesis. Plant Cell *19*,
606   2039–2052.

607 34. Goossens, J., Swinnen, G., Vanden Bossche, R., Pauwels, L., and Goossens, A. (2015).
608   Change of a conserved amino acid in the MYC2 and MYC3 transcription factors leads to
609   release of JAZ repression and increased activity. New Phytol. *206*, 1229–1237.

610 35. Mine, A., Seyfferth, C., Kracher, B., Berens, M. L., Becker, D., and Tsuda, K. (2018).
611   The Defense Phytohormone Signaling Network Enables Rapid, High-Amplitude
612   Transcriptional Reprogramming during Effector-Triggered Immunity. Plant Cell *30*,
613   1199–1219.

614 36. Innes, R. W., Bisgrove, S. R., Smith, N. M., Bent, A. F., Staskawicz, B. J., and Liu, Y. C.
615   (1993). Identification of a disease resistance locus in Arabidopsis that is functionally
616   homologous to the RPG1 locus of soybean. Plant J. *4*, 813–820.

617 37. Ryu, K. H., Huang, L., Kang, H. M., and Schiefelbein, J. (2019). Single-Cell RNA
618   Sequencing Resolves Molecular Relationships Among Individual Plant Cells. Plant
619   Physiol. *179*, 1444–1456.

620 38. Shulse, C. N., Cole, B. J., Ciobanu, D., Lin, J., Yoshinaga, Y., Gouran, M., Turco, G. M.,
621   Zhu, Y., O'Malley, R. C., Brady, S. M., et al. (2019). High-Throughput Single-Cell
622   Transcriptome Profiling of Plant Cell Types. Cell Rep. *27*, 2241–2247.e4.

623 39. Biswas, S., Kerner, K., Teixeira, P. J. P. L., Dangl, J. L., Jojic, V., and Wigge, P. A.
624   (2017). Tradict enables accurate prediction of eukaryotic transcriptional states from 100
625   marker genes. Nat. Commun. *8*, 15309.

626 40. Park, P. J. (2009). ChIP-seq: advantages and challenges of a maturing technology. Nat.
627   Rev. Genet. *10*, 669–680.

628 41. Buenrostro, J. D., Wu, B., Chang, H. Y., and Greenleaf, W. J. (2015). ATAC-seq: A
629   Method for Assaying Chromatin Accessibility Genome-Wide. Curr. Protoc. Mol. Biol. *109*,
630   21.29.1-9.

631 42. Falk, A., Feys, B. J., Frost, L. N., Jones, J. D., Daniels, M. J., and Parker, J. E. (1999).
632   EDS1, an essential component of R gene-mediated disease resistance in Arabidopsis
633   has homology to eukaryotic lipases. Proc. Natl. Acad. Sci. USA *96*, 3292–3297.

634 43. Gallego-Giraldo, L., Escamilla-Trevino, L., Jackson, L. A., and Dixon, R. A. (2011).
635   Salicylic acid mediates the reduced growth of lignin down-regulated plants. Proc. Natl.
636   Acad. Sci. USA *108*, 20814–20819.

637 44. Okayama, H., and Berg, P. (1982). High-efficiency cloning of full-length cDNA. Mol. Cell.
638   Biol. *2*, 161–170.

639 45. Rallapalli, G., Kemen, E. M., Robert-Seilaniantz, A., Segonzac, C., Etherington, G. J.,
640   Sohn, K. H., MacLean, D., and Jones, J. D. G. (2014). EXPRSS: an Illumina based
641   high-throughput expression-profiling method to reveal transcriptional dynamics. BMC
642   Genomics *15*, 341.

643 46. Picelli, S., Björklund, A. K., Reinius, B., Sagasser, S., Winberg, G., and Sandberg, R.
644   (2014). Tn5 transposase and tagmentation procedures for massively scaled sequencing
645   projects. Genome Res. *24*, 2033–2040.

646 47. Swarbreck, D., Wilks, C., Lamesch, P., Berardini, T. Z., Garcia-Hernandez, M., Foerster,
647   H., Li, D., Meyer, T., Muller, R., Ploetz, L., et al. (2008). The Arabidopsis Information
648   Resource (TAIR): gene structure and function annotation. Nucleic Acids Res. *36*,
649   D1009-14.

650 48. Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990). Basic local
651   alignment search tool. J. Mol. Biol. *215*, 403–410.

652 49. Trapnell, C., Roberts, A., Goff, L., Pertea, G., Kim, D., Kelley, D. R., Pimentel, H.,
653   Salzberg, S. L., Rinn, J. L., and Pachter, L. (2012). Differential gene and transcript
654   expression analysis of RNA-seq experiments with TopHat and Cufflinks. Nat. Protoc. *7*,
655   562–578.

656 50. Langmead, B., and Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2.
657   Nat. Methods *9*, 357–359.

658 51. Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G.,
659   Abecasis, G., Durbin, R., and 1000 Genome Project Data Processing Subgroup (2009).
660   The Sequence Alignment/Map format and SAMtools. Bioinformatics *25*, 2078–2079.

661
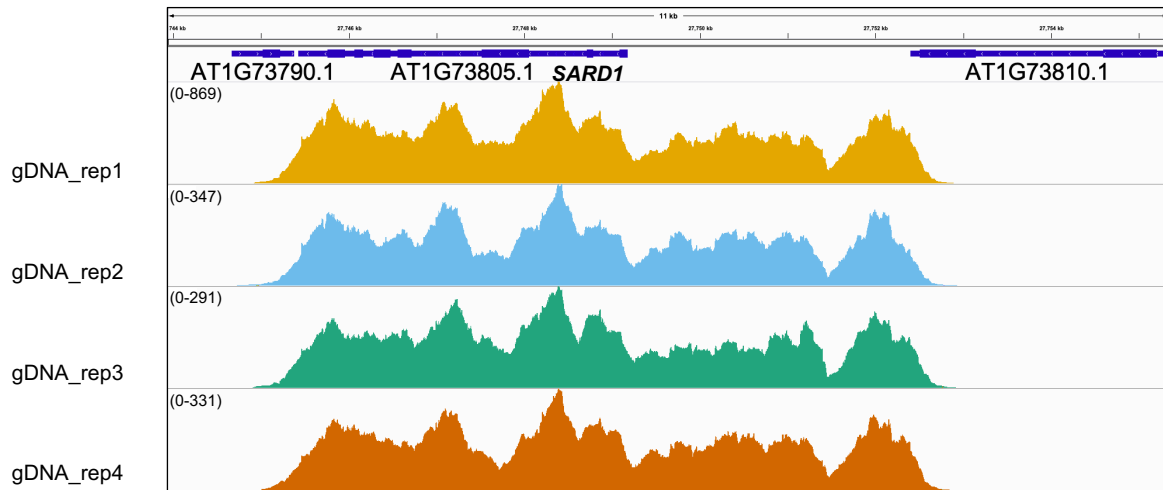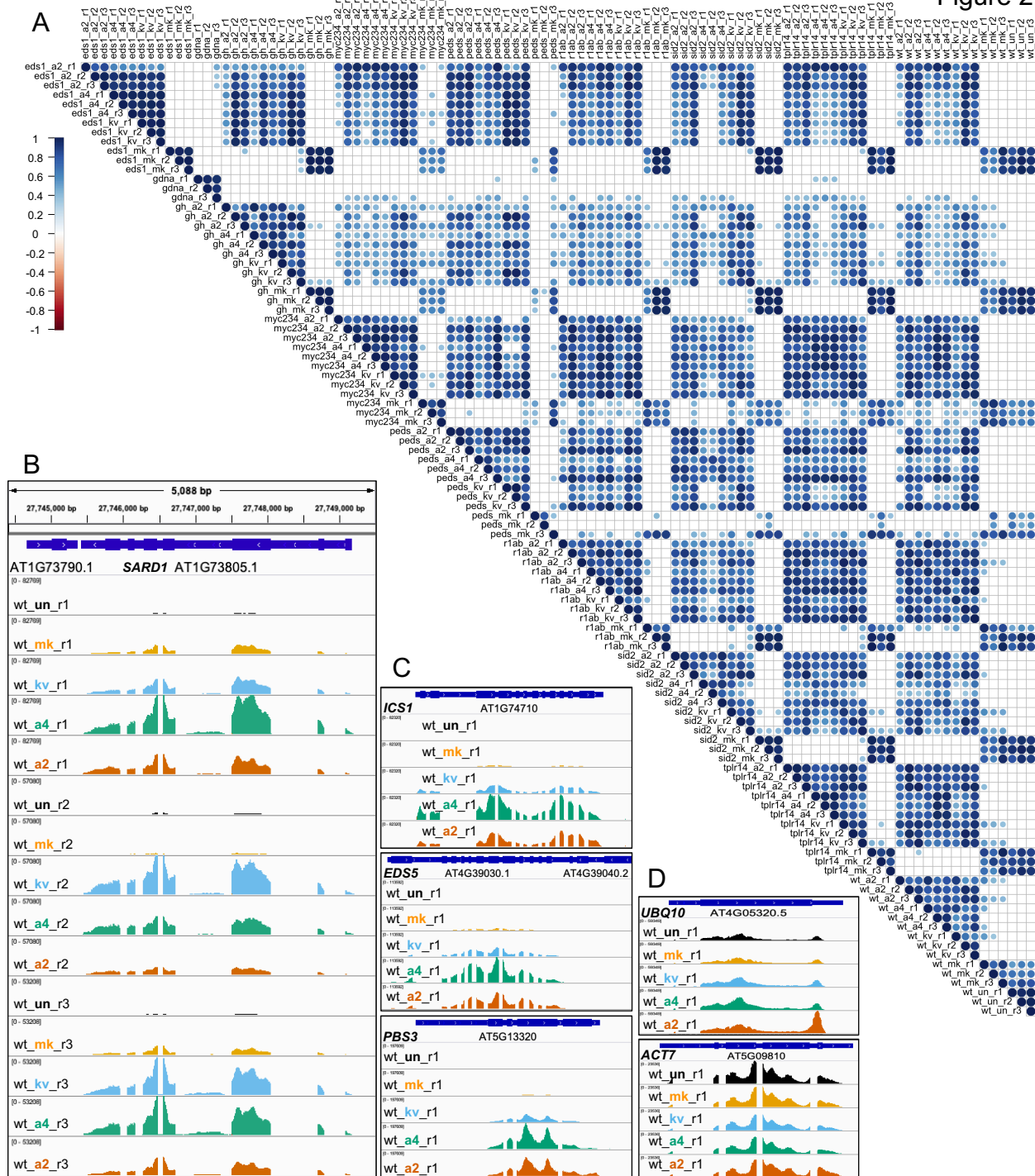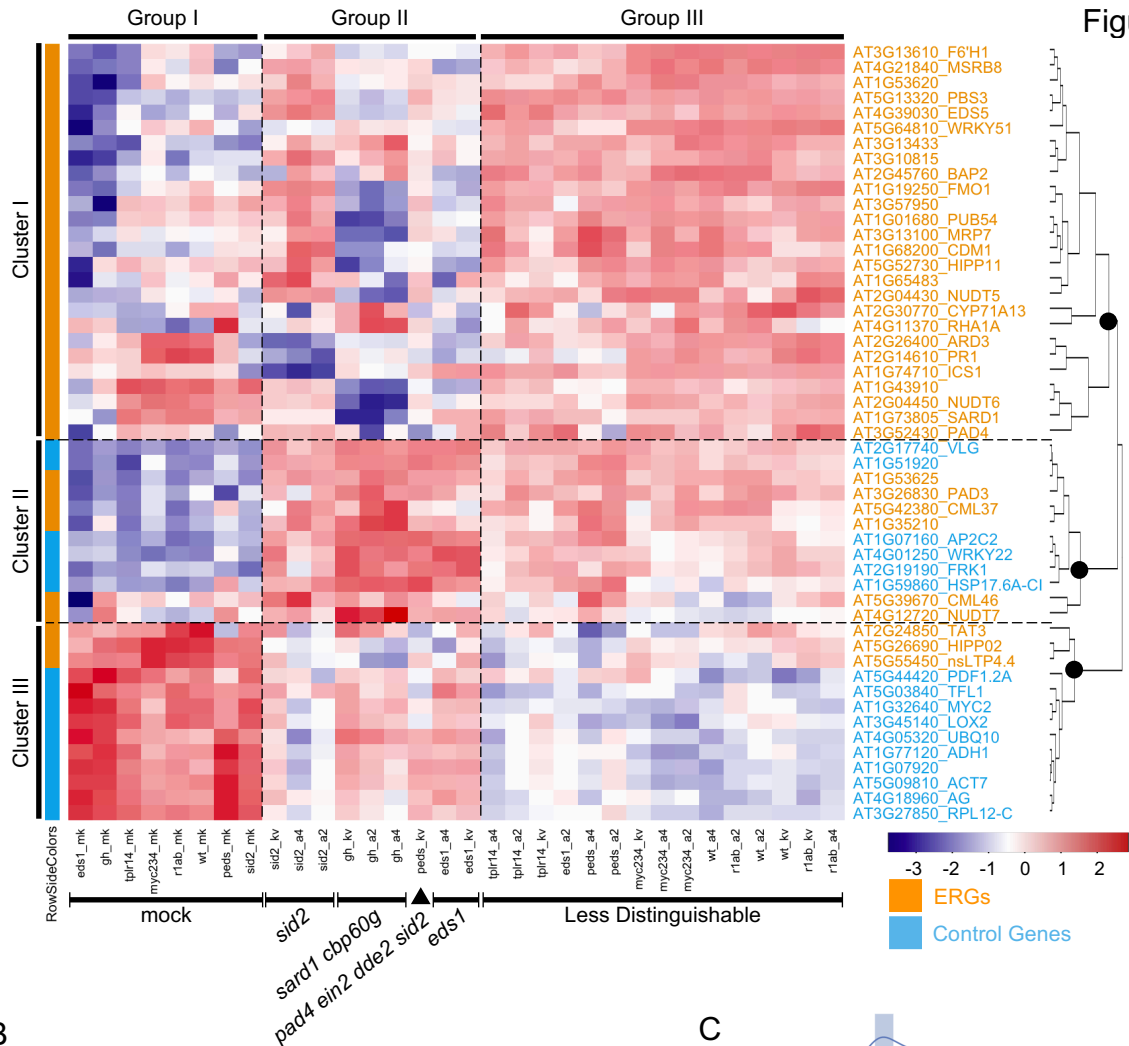
Figure 1

A

B

Figure 2

Figure 3

A — Heatmap with Group I, Group II, Group III and Cluster I, Cluster II, Cluster III.

Gene labels (top to bottom):
AT3G13610_F6'H1, AT4G21840_MSRB8, AT1G53620, AT5G13320_PBS3, AT4G39030_EDS5, AT5G64810_WRKY51, AT3G13433, AT3G10815, AT2G45760_BAP2, AT1G19250_FMO1, AT3G57950, AT1G01680_PUB54, AT3G13100_MRP7, AT1G68200_CDM1, AT5G52730_HIPP11, AT1G65483, AT2G04430_NUDT5, AT2G30770_CYP71A13, AT4G11370_RHA1A, AT2G26400_ARD3, AT2G14610_PR1, AT1G74710_ICS1, AT1G43910, AT2G04450_NUDT6, AT1G73805_SARD1, AT3G52430_PAD4, AT2G17740_VLG, AT1G51920, AT1G53625, AT3G26830_PAD3, AT5G42380_CML37, AT1G35210, AT1G07160_AP2C2, AT4G01250_WRKY22, AT2G19190_FRK1, AT1G59860_HSP17.6A-CI, AT5G39670_CML46, AT4G12720_NUDT7, AT2G24850_TAT3, AT5G26690_HIPP02, AT5G55450_nsLTP4.4, AT5G44420_PDF1.2A, AT5G03840_TFL1, AT1G32640_MYC2, AT3G45140_LOX2, AT4G05320_UBQ10, AT1G77120_ADH1, AT1G07920, AT5G09810_ACT7, AT4G18960_AG, AT3G27850_RPL12-C

Column labels: eds1_mk, gh_mk, tplr14_mk, myc234_mk, r1ab_mk, wt_mk, peds_mk, sid2_mk, sid2_a4, sid2_a2, gh_kv, gh_a2, gh_a4, peds_kv, eds1_a4, eds1_kv, tplr14_a4, tplr14_a2, tplr14_kv, eds1_a2, peds_a4, peds_a2, myc234_kv, myc234_a4, myc234_a2, wt_a4, r1ab_a2, wt_a2, wt_kv, r1ab_kv, r1ab_a4

mock; sid2; sard1 cbp60g; pad4 ein2 dde2 sid2; eds1; Less Distinguishable

ERGs (orange); Control Genes (blue)

B

	GO term name (GO:BP, KEGG, WP)	p-value
Cluster I	systemic acquired resistance	7.69E-08
	salicylic acid biosynthetic process	1.72E-06
	defense response, incompatible interaction	1.71E-05
	defense response to bacterium	8.90E-05
	programmed cell death	3E-04
Cluster II	MAPK signaling pathway - plant	0.0040897
	plant-pathogen interaction	0.0082958
	response to reactive oxygen species	0.0453463
	dihydrocamalexic acid decarboxylase activity	0.04994
Cluster III	response to wounding	2.90E-05
	flower development	0.001508
	response to jasmonic acid	0.003898
	tyrosine metabolism	0.027459
	alpha-linolenic acid metabolism	0.033211

C — Scatter plot: x-axis wt_a2 (PTI + cETI), y-axis wt_a4(PTI + tETI), $R^2=0.81$

Figure 4

Supplementary Figure S1

A

52 gene loci selected
261,616 bp genomic region

target selection
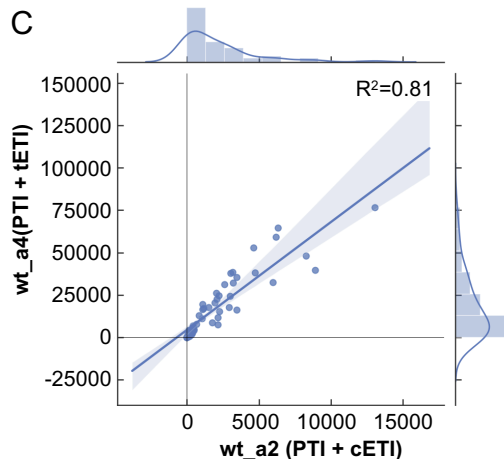
Scripts (Github)
"RepeatMasker" ■
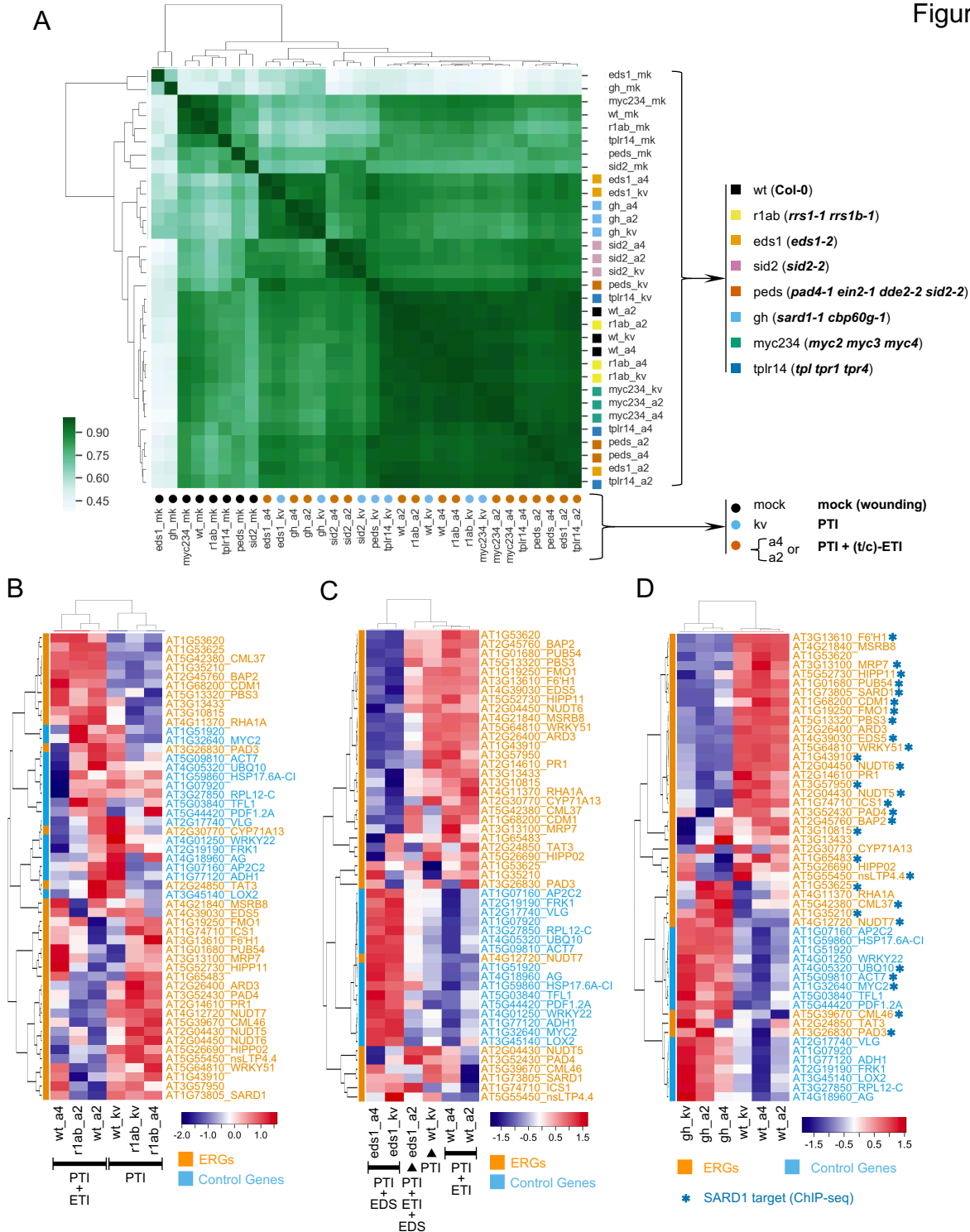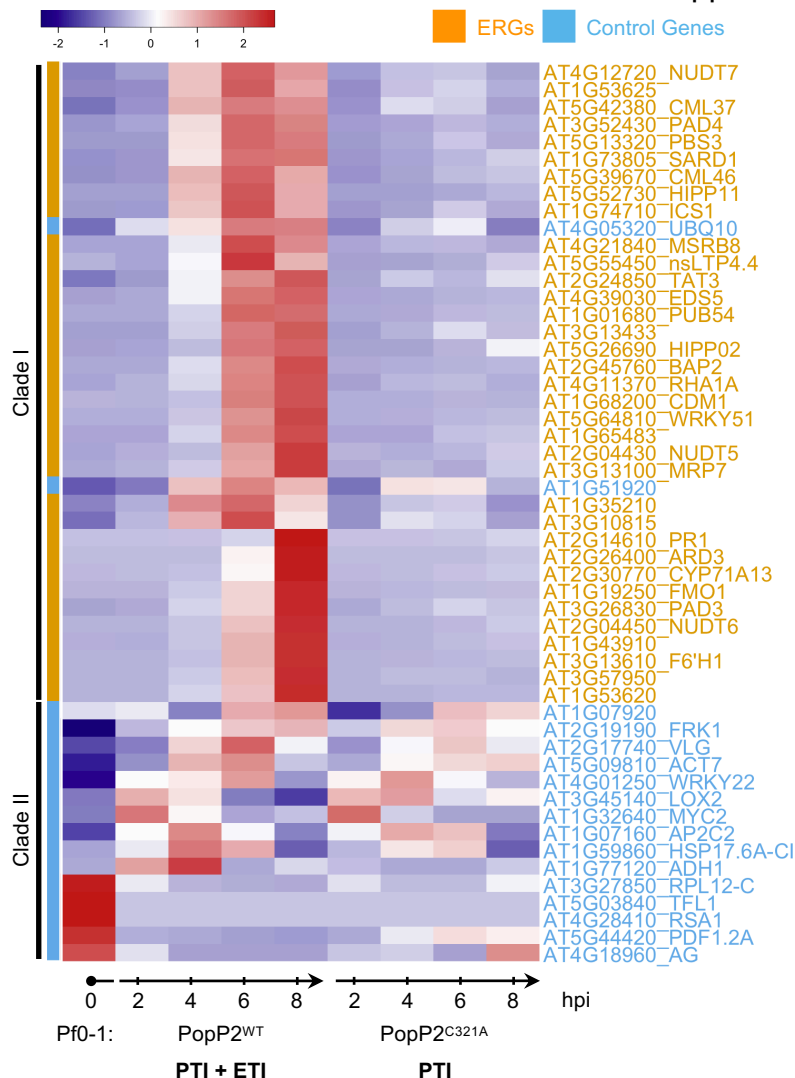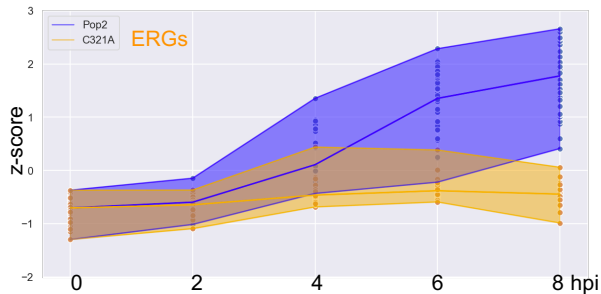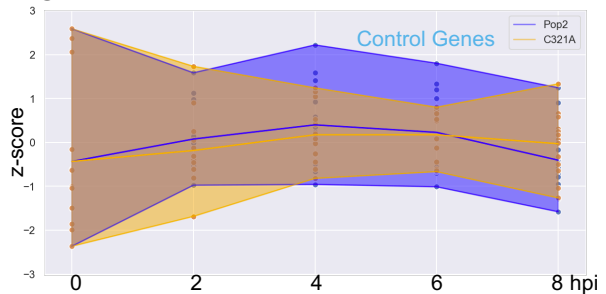
FASTA files (Github)
◆ **ETI_capture_gene_list.fa**

18,800 bp masked

masking repeats

→ ETI_capture_gene_list.masked.processed.fa

2221×120-mer ssRNAs
with 17-bp tiling

initial bait library

"FastaParser_makeBaitLibs_fromFasta.pl" ■

→ ETI_capture_gene_lib.fa ◆

MEGABLAST baits to wild-
type reference genome

reduction of
bait redundancy I

"MEGABLAST" ■

→ ETI_capture_gene_lib_120_17.fa

MEGABLAST baits
to each other

reduction of
bait redundancy II

"MEGABLAST" ■

final bait library

→ **ETI_capture_gene_lib_120_17_final.fa**

**2219 × 120-mer
ssRNA baits**

B

| Gene of interest | AT1G53620 | AT1G53625 |
| AT1G53610 | | AT1G53630 |

Coding strand

Bait region

Final baits:
120 nt/each

Coordinates

20.009M   20.01M   20.011M   20.012M   20.013M   20.014M   20.015M   20.016M   20.017M

C

	10 mM MgCl$_2$	mock	**mk**
	Pf0-1:AvrRps4^{mut}	PTI	**kv**
	Pf0-1:AvrRps4	PTI + t-ETI	**a4**
	Pf0-1:AvrRpt2	PTI + c-ETI	**a2**
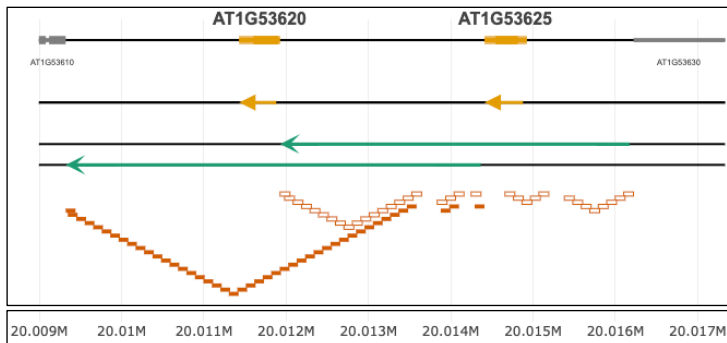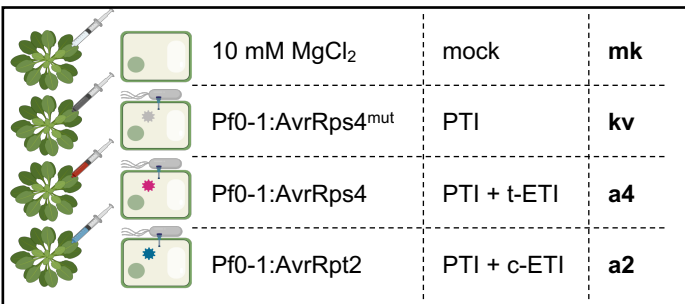
4 × 8 = **32**

Untreated (**unt**) | Col-0 (**wt**) **1**

Conditions: 1 + 32 = 33   Biological replicates (r1~r3): 3×

Total number of RNA-seq libraries : 33 × 3 = 99

WT	Col-0	Ref.	**wt**
mutants	*eds1-2*	[1]	**eds1**
	rrs1-3 rrs1b-1	[2]	**r1ab**
	sard1-1 cbp60g-1	[3]	**gh**
	myc2 myc3 myc4	[4]	**myc234**
	tpr1 tpl tpr4	[5]	**tplr14**
	sid2-2	[6]	**sid2**
	pad4-1 ein2-2 dde2-1 sid2-2	[7]	**peds**

Supplementary Figure S4