Genomic basis of transcriptome dynamics in rice under field conditions

Makoto Kashima[1], Ryota L. Sakamoto[2], Hiroki Saito[3,4], Satoshi Ohkubo[3,5], Ayumi Tezuka[1], Ayumi Deguchi[1], Yoichi Hashida[1], Yuko Kurita[1], Koji Iwayama[5], Shunsuke Adachi[5] and Atsushi J. Nagano[7]*

[1]Research Institute for Food and Agriculture, Ryukoku University, Shiga, Japan

[2]Seibi Senior High School, Gifu, Japan

[3]Graduate School of Agriculture, Kyoto University, Kyoto, Japan

[4]Tropical Agriculture Research Front, Japan International Research Center for Agricultural Sciences, Okinawa, Japan

[5] Faculty of Data Science, Shiga University, Shiga, Japan

[6] Institute of Global Innovation Research, Tokyo University of Agriculture and Technology, Fuchu, Tokyo, Japan

[7]Faculty of Agriculture, Ryukoku University, Shiga, Japan

*e-mail: anagano@agr.ryukoku.ac.jp

How genetic variations affect gene expression dynamics of field-grown plants remains unclear. Using statistical analysis of large-scale time-series RNA-sequencing of field-grown rice from chromosome segment substitution lines (CSSLs), we identified 1675 expression dynamics quantitative trait loci (edQTLs) leading to polymorphisms in expression dynamics under field conditions. Based on the edQTL and environmental information, we successfully predicted gene expression under environments different from training environments, and in rice cultivars with more complex genotypes than the CSSLs. Overall, edQTL' identification helped understanding the genetic architecture of expression dynamics under field conditions, which is difficult to assess with laboratory experiments[1].The prediction of expression based on edQTL and environmental information will contribute to crop breeding by increasing the accuracy of trait prediction under diverse conditions.

Organisms respond to fluctuations in field environments variably depending on their genetic backgrounds, developmental stages, and physiological status. This variability can cause missing heritability in crop breeding and low accuracy in medicine[2–4]. Environmental stimuli can induce transcriptional responses directly and/or indirectly. Measuring transcriptome dynamics is a comprehensive method for assessing environmental responses and their alteration due to genetic, developmental, and

36 physiological factors. The expression quantitative trait loci (eQTL) approach is frequently
37 used to assess the association between genetic variation and gene expression
38 polymorphism[5–9]. Because this approach requires transcriptome data for the complete set
39 of a QTL mapping population under the given conditions[5–9], only a limited range of
40 environmental conditions are covered. Contrarily, statistical models based on
41 meteorological information, circadian clock, and developmental age have succeeded in
42 describing transcriptome dynamics in the field[10–12], although these models can only
43 predict the transcriptome of one or few genotypes used as training data[10–12]. Thus, no
44 study has clarified the relationship between genetic variation and transcriptome dynamics
45 under fluctuating field conditions.

46 Here, we first defined the loci that determine polymorphisms in gene expression
47 dynamics as 'edQTL'. edQTL can explain polymorphisms in gene expression dynamics
48 under a broad range of environmental conditions, while eQTL can only explain them
49 under the investigated environment (Fig. 1a). Our approach consists of four steps: time-
50 series RNA-sequencing (RNA-Seq), prediction model development, edQTL detection,
51 and evaluation of edQTL (See Methods, Fig. 1b, and Supplementary Fig. 1). We used the
52 rice (*Oryza sativa* L.) cultivars 'Koshihikari' (a leading *japonica* cultivar) and 'Takanari'
53 (a high-yield *indica* cultivar)[13,14] that would present substantial polymorphisms in

54 expression dynamics. The 78 reciprocal CSSLs and two backcross inbred lines (BILs)
55 developed from the parental lines were also used[14–16] (Fig. 1b and Supplementary Fig.
56 2a). To refelct the effect of plant age (days after seeding) in our model, we prepared four
57 different sets of rice plants transplanted at two-weeks intervals. Sixteen sets (days) of
58 bihourly sampling for 24 h were conducted from May to September 2015 (cropping
59 season in Japan). At each sampling time, the youngest fully expanded leaves were
60 collected from two different genotypes at each transplant set while withered plants were
61 excluded from the sampling, which could not be sampled (Fig. 1c, Supplementary Fig.
62 2b, and Table S2). The 926 individual leaves were applied for RNA-Seq, to obtain
63 transcriptome data. After filtering samples and genes (Supplementary Fig. 3a, b) and
64 confirming genotypes (Supplementary Fig. 3c and 4), 23,924 expressed genes from 854
65 samples were used. Correlation analysis of the transcriptome data showed an obvious
66 transcriptome-wide diurnal variation in the expressed genes (Fig. 1d).

67 We then developed two prediction models describing transcriptome dynamics under
68 fluctuating field environments in 'Koshihikari' and 'Takanari'. In this step, we used the
69 statistical modeling tool 'FIT'[10] for predicting transcriptome dynamics under field
70 conditions. As most of the genome in the CSSLs was not substituted (Supplementary Fig
71 2a), the expression dynamics of most genes was expected to be identical to that of the

72 background parent. Thus, the RNA-Seq data of 'Koshihikari'- and 'Takanari'-background
73 samples were used to develop the prediction model for 'Koshihikari' and 'Takanari' lines,
74 respectively. Circadian clock and meteorological data (air temperature and global solar
75 radiation) (Supplementary Fig. 5a) were considered in model development and "scaled
76 age" was used to adjust differences in heading date among rice genotypes (See Methods
77 and Supplementary Fig. 5). We found polymorphisms in the predicted expression
78 dynamics of 3696 genes (15.4% of the expressed genes; Supplementary Fig. 3e and Table
79 S3). For instance, the 'Koshihikari' model for *Os09g0343200* showed an obvious diurnal
80 oscillation in expression, whereas the 'Takanari'-model showed a constant low-level
81 expression (Fig. 2a). Notably, in some 'Takanari'-background CSSLs, SL1329 and
82 SL1330, the expression of *Os09g0343200* resembled that of the 'Koshihikari' model
83 rather than the 'Takanari'-model (Fig. 2a), suggesting that genetic substitution in the
84 chromosome 9 affected *Os09g0343200* expression.

85 The CSSLs had been previously genotyped using 141 simple sequence repeat (SSR)
86 markers (Supplementary Fig. 2a)[14]. As so, at the edQTL detection step, we searched for
87 the SSR markers explaining the expression dynamics polymorphisms. We evaluated the
88 decrease in the residual error of each gene assuming edQTL around each SSR marker
89 (See Methods and Supplementary Fig. 6a). For *Os09g0343200*, the sum of residual errors

90 significantly decreased only by assuming edQTL on chromosome 9 (Fig. 2b and
91 Supplementary Fig. 6b), indicating that this is a *cis*-edQTL. The prediction models chosen
92 for each sample based on edQTL genotypes explained the expression dynamics of
93 *Os09g0343200* better than their background genotypes (Fig. 2c). Overall, edQTL were
94 identified for 1675 genes (45.3% of genes with expression dynamics polymorphism
95 between 'Koshihikari' and 'Takanari', false discovery rate = 0.05; Supplementary Fig.
96 6f), including 222 genes affected by *trans*-edQTL (Fig. 2d). Forty-three genes were
97 affected by multiple edQTL. For 33 of these genes, the sum of residual errors based on
98 all edQTL was smaller than that of the most significant edQTL alone (Supplementary Fig.
99 7). Detailed results for individual genes can be found in our database
100 (https://ps.agr.ryukoku.ac.jp/osa_edqtl). A *cis*-edQTL could explain the expression
101 dynamics polymorphism of *Os01g0537250* that be specifically observed in young plants
102 (Fig. 2e,f). The expression depended on time of day in both parental models, but scaled
103 age was only important in the 'Koshihikari' model (Fig. 2e,f and Supplementary Fig. 8).
104 The edQTL-affected 393 genes depending on the time of day and 275 genes depending
105 on scaled age (Supplementary Fig. 9).

106 In the edQTL evaluation step, we first compared the predicted performances based
107 on the edQTL model and background genotypes model (BG model) under different

108 environments from where training data were obtained for model development.
109 'Koshihikari', 'Takanari', and the CSSLs of two transplant sets were cultivated in a field
110 different from that used in 2015 and sampled in August 2016; 139 RNA-Seq datasets were
111 obtained (Supplementary Fig. 2c and Table S2). Environmental factors differed between
112 the 2015 and 2016 fields (Supplementary Fig. 10a). For 91.8% of the 1675 genes affected
113 by the edQTL, the prediction was improved by the edQTL model compared with the BG
114 model (Fig. 3a,b and Supplementary Fig. 10b). Furthermore, the prediction of the
115 expressed genes for the validation dataset showed comparable accuracy to that of the
116 training dataset (Supplementary Fig. 11). Thus, the identified edQTL could explain the
117 expression dynamics polymorphisms in different years and locations. For further
118 verification , two BILs between 'Koshihikari' and 'Takanari' (HP-a and HP-b)[15,16] were
119 cultivated at Takatsuki in 2015, and five and six RNA-Seq datasets were obtained for HP-
120 a and HP-b, respectively (Supplementary Fig 2b and Table S2). These lines carried
121 'Koshihikari' alleles at 16.3% and 19.9% of their markers with the genetic background of
122 'Takanari' (Supplementary Fig 2a). To evaluate the performance of the edQTL-based
123 prediction, we calculated the sum of prediction errors of all edQTL-influenced genes
124 using edQTL-, 'Koshihikari'-, and 'Takanari'-models. Overall, the edQTL model
125 provided the best prediction (Fig. 3c,d). Permutation analysis of markers in HP-a and HP-

126 b genomes showed the significant advantage of the edQTL model ($p < 0.001$) (Fig. 3c,d).
127 For instance, as the genotype of the *trans*-edQTL for *Os03g0388300* comprised the
128 'Koshihikari' allele in the genome of HP-a and the 'Takanari' allele in the genome of HP-
129 b (Fig. 3e), 'Koshihikari'- and 'Takanari'-models were used to predict *Os03g0388300*
130 expression in the edQTL models for HP-a and HP-b, respectively. The expression of
131 *Os03g0388300* fluctuated with time and it was higher in 'Koshihikari' than in 'Takanari'
132 (Supplementary Fig. 12a). The prediction of *Os03g0388300* expression based on the
133 edQTL models was better than the prediction based on the BG models (Fig. 3f).
134 Regarding *OsKS3* (*Os04g0611700*)[17] in HP-b, the prediction based on the edQTL model
135 was worse than that based on the BG model (Supplementary Fig. 12b,c). True edQTL
136 might therefore exist around SSR markers 52 or 53, which were not substituted with the
137 'Koshihikari' allele in HP-b (Supplementary Fig. 12b). Such difficulty may occur in some
138 edQTL-based predictions for the BILs, because several edQTLs identified by the CSSLs
139 can be unlinked to the substituted genome regions in the BILs. Finally, we concluded that
140 our approach successfully identified loci linked to expression dynamics polymorphisms
141 in field conditions.

142 The edQTL approach scans a broader range of conditions (Fig. 1a) but it is less
143 sensitive when focusing on a specific condition. This is because eQTL are expected to

144 show the same effects on gene expressions in all samples used in a study. Contrarily, some
145 edQTL were expected to affect gene expressions in some specific samples, depending on
146 their environments. The lower sensitivity of the edQTL approach relative to that of the
147 eQTL approach might explain the smaller fraction of *trans*-edQTL compared with that of
148 *trans*-eQTL in previous studies (13.3% and 62–71%[6,8,9], respectively) because *trans*-
149 edQTL/eQTL have generally smaller effect sizes than *cis*-edQTL/eQTL[6,8].

150 Phenotypic plasticity plays key roles in plant's environmental adaptation[18,19].
151 Nevertheless, as this is difficult to study in natural environments, the genetic architecture
152 of phenotypic plasticity under field conditions remains largely unveiled. The edQTL
153 identified by our novel method improve the understanding of the genetic architecture
154 underlying the expression dynamics polymorphisms between 'Koshihikari' and 'Takanari'
155 in the field. In addition, our method showed that rice gene expression dynamics can be
156 predicted based on genotypic and meteorological information. In crop breeding,
157 polymorphisms of environmental responses result in unexpected performance of a bred
158 crop under environments differing from experimental fields. Because the transcriptome
159 can be beneficial for trait prediction[20–23] and our approach allows transcriptome
160 predictions under various conditions, it contributes for crop breeding and for
161 understanding plant systems.

162

163 METHODS

164 **Overview of edQTL identification and verification.** Our approach consisted of four
165 steps (Fig. 1b). First, the parent rice lines and their descendants were cultivated in a paddy
166 field and sparsely sampled at several time points for RNA-Seq (Time-series RNA-Seq,
167 Fig. 1b and Supplementary Fig. 1). Second, parental prediction models were developed
168 to describe environmental responses in 'Koshihikari' and 'Takanari' in terms of gene
169 expression (Prediction model development, Fig. 1b). Most gene expression dynamics
170 would be identical among the CSSLs and the parent with the same background genotype.
171 Thus, we used 'Koshihikari'- and 'Takanari'-background CSSLs as well as their parents
172 to develop the parental models (Fig. 1b and Supplementary Fig. 1). Third, the dependency
173 of expression dynamics polymorphisms on genetic variation was statistically evaluated
174 based on comparisons between predictive gene expression and observed gene expression
175 in CSSLs assuming that genetic variation of SSR markers leads to expression dynamics
176 polymorphisms (Supplementary Fig. 6) (edQTL detection, Fig. 1b). Fourth, by integrating
177 the parental models and edQTL information, gene expression dynamics was predicted
178 based on environmental and genotypic information (Evaluation of edQTL).

179 **Plant materials**. We cultivated the following rice (*O. sativa*) lines: the *japonica* variety
180 'Koshihikari', the *indica* variety 'Takanari', 78 reciprocal CSSLs (40 'Koshihikari'-

181 background lines, except for SL1213, and 38 'Takanari'-background lines, except for
182 SL1306)[15], and two 'Takanari'-background BILs (HP-a and HP-b)[15,16]. Each variety and
183 line were sown in nursery trays. Approximately one month after sowing, seedlings were
184 transplanted to a paddy field at Takatsuki, Japan (34°51′19″N, 135°37′51″E) in 2015 and
185 at Kizugawa, Japan (34°4′05″N, 135°50′33″E) in 2016. To consider the effect of plant age
186 in the prediction of gene expression dynamics, four and two transplant sets were prepared
187 in 2015 and 2016, respectively (Fig. 1c and Supplementary Fig. 2b,c). Seed sowing and
188 transplanting were conducted according to the following schedules. In 2015, transplant
189 set 1: Seed sowing date "2015-04-03", Transplanting date "2015-05-01"; transplant set 2:
190 Seed sowing date "2015-04-17", Transplanting date "2015-05-08"; transplant set 3: Seed
191 sowing date "2015-05-01", Transplanting date "2015-05-22"; transplant set 4: Seed
192 sowing date "2015-05-15", Transplanting date "2015-06-05". In 2016, transplant set 1:
193 Seed sowing date "2016-04-21", Transplanting date "2016-05-16"; transplant set 2: Seed
194 sowing date "2016-05-19", Transplanting date "2016-06-08".

195 **Sampling and RNA extraction**. Sixteen sets (2015) and three sets (2016) of bihourly
196 sampling for 22 h, from 16:00 on one day to 14:00 on the next, were conducted on the
197 following dates (Supplementary Fig. 2b,c and Table S2): 5–6 May, 16–17 June, 22–23
198 June, 29–30 June, 6–7 July, 13–14 July, 20–21 July, 27–28 July, 3–4 August, 19–20

199 August, 24–25 August, 31 August to 1 September, 7–8 September, 14–15 September, 21–
200 22 September, and 28–29 September in 2015; and 4–5, 11–12, and 25–26 August in 2016.
201 We applied a stratified randomization strategy to the sampling schedule to avoid biased
202 sampling of each line over seasons. We separated individual plants into four and two
203 groups in each transplant set in 2015 and 2016, respectively, each containing the same
204 number of individuals per line. Then, the order of sampling was randomized in each group.
205 According to the sampling schedule, two plants from the 82 genotypes of each transplant
206 set were sampled. Because aged rice withered, several samples were missed at the latest
207 cultivation (Supplementary Fig. 2b). Transplant set and sampling time for each sample
208 are listed in Table S2. The youngest fully expanded leaf from each plant was collected,
209 immediately frozen in liquid nitrogen, and stored at -80 °C until RNA isolation for RNA-
210 Seq. Individual plants were only sampled once to avoid wounding response. Thus, all
211 RNA-Seq data were obtained from independent plants. Frozen samples were
212 homogenized with TissueLyser II (Qiagen, Hilden, Germany), and total RNA was then
213 extracted using the Maxwell 16 LEV Plant RNA Kit (Promega, Madison, WI, USA) and
214 Maxwell 16 Automated Purification System (Promega). The concentration of RNA was
215 measured using the Quant-iT RNA Assay Kit, broad range (Thermo Fisher Scientific,
216 Waltham, MA, USA).

RNA-Seq library preparation and sequencing. An automated liquid handling system Freedom EVO 150 (TECAN, Zurich, Switzerland) and a thermal cycler ODTC 384 (INHECO, Martinsried, Germany) were utilized to prepare the RNA-Seq library of 384 samples simultaneously. The protocol of RNA-Seq library preparation is described below. We enzymatically degraded abundant RNAs such as rRNAs in the leaves, which cause wasteful consumption of sequence reads (Supplementary Fig. 13). Aliquots of 252 and 192 types of 100-μM 60-mer antisense DNAs (IDT, San Jose, CA, USA) covering rRNAs and the top eight most frequently-detected transcripts (*Osp1g00110.1*, *Osp1g00180.1*, *Osp1g00420.1*, *Osp1g00170.1*, *Osp1g00340.1*, *Osp1g00330.1*, *Osp1g00600.1*, and *Os11t0707000-02*) in the RNA-Seq dataset of *O. sativa* leaf were mixed in equal amounts (SDRNA oligo pool)[24,25]. One microgram of total RNA, 1.5 μL of the SDRNA oligo pool, 3 μL of 5× hybridization buffer (0.5 M Tris-HCl (pH 7.4) and 1 M NaCl), and RNase-free water were mixed to a total volume of 10 μL. The mixture was incubated to anneal the antisense DNAs with the target transcripts using the thermal cycler ODTC 384 with the following program: 95 °C for 2 min, slow cooling to 45 °C (0.1 °C/s), and 45 °C for 5 min. Subsequently, 1 μL of Tli Thermostable RNase H (TaKaRa, Kusatsu, Japan), 2 μL of 10× RNase H digestion buffer [500 mM Tris-HCl (pH 7.4), 1 M NaCl, and 200 mM MgCl], and 7 μL of RNase-free water were added and mixed well. This mixture was

235 incubated at 45 °C for 30 min to selectively digest the RNA of the DNA/RNA hybrid
236 double strand. Then, 24 μL of AMPure XP (Beckman Coulter, Brea, CA, USA) beads
237 was added and the mixture was purified in a 384-well magnetic plate following the
238 manufacturer's instructions. The RNA was then eluted with 14 μL of RNase-free water.
239 Ten microliters of the RNA solution, 1 μL of DNaseI (5 unit/μL), 1 μL of 10× DNaseI
240 buffer (both Promega), 1 μL of 100 mM DTT (Invitrogen, Carlsbad, CA, USA), and 7 μL
241 of RNase-free water were mixed well. This mixture was incubated at 37 °C for 30 min to
242 degrade the SDRNA oligo pool. Subsequently, 24 μL of AMPure XP beads was added
243 and purified according to the manufacturer's instructions. The RNA was eluted with 14
244 μL of RNase-free water. Five microliters of the purified mRNA obtained was mixed with
245 4 μL of 5× SS buffer (Invitrogen), and 1 μL of 100 mM DTT. Fragmentation of mRNA
246 was carried out at 94 °C for 4.5 min and immediately cooled on ice. Subsequently, 0.6
247 μL of 100 μM random primer $(N)_6$ (Promega) and 0.9 μL of RNase-free water were added
248 to the mixture, which was incubated at 50 °C for 5 min and immediately chilled on ice to
249 relax the secondary structure of the RNA. The fragmented RNA with random hexamers
250 and the reverse transcription master mix [1 μL of 100 mM DTT, 0.4 μL of dNTPs (25
251 mM each; Promega), 0.1 μL of SuperScript IV (Invitrogen), 0.2 μL of actinomycin D
252 (1000 ng/μL) (Nacalai Tesque, Kyoto, Japan), and 5.9 μL of RNase-free water] were

253 mixed. For the reverse transcription step, the mixture was incubated at 25 °C for 10 min,
254 followed by 50 min at 50 °C. SuperScript IV was inactivated by heating the mixture at
255 75 °C for 15 min. Subsequently, 24 μL of AMPure XP and 12 μL of 99.5% ethanol were
256 added, and the mixture was purified according to the manufacturer's protocol. The reverse
257 transcription product was eluted with 14 μL of RNase-free water. The purified DNA/RNA
258 hybrid solution without beads and the second-strand synthesis master mix [2 μL of 10×
259 Blue Buffer (Enzymatics, Beverly, MA, USA), 1 μL of dUTP/NTP mix (Fermentas,
260 Burlington, Canada), 0.5 μL of 100 mM DTT, 0.5 μL of RNase H (Enzymatics), 1 μL of
261 DNA polymerase I (Enzymatics), and 5 μL of RNase-free water] were mixed. This
262 mixture was incubated at 16 °C for 4 h, and then purified with 24 μL of AMPure XP
263 according the manufacturer's protocol. The purified dsDNA was eluted with 14 μL of
264 RNase-free water. Five microliters of the dsDNA solution was used in the following step.
265 End-repair, A-tailing, and adapter ligation were carried out using the KAPA Hyper prep
266 kit (KAPA BIOSYSTEMS, Wilmington, MA, USA) with 1/10× volume of the solutions
267 according the manufacturer's protocol. One microliter of 0.1 μM Y-shape adapter[24] was
268 used in the adapter ligation step for 15 min. Then, size selection of the ligation product
269 was performed with 5.5 μL of AMPure XP. The purified dsDNA was eluted with 10 μL
270 of RNase-free water and a second round of size selection was performed with 10 μL of

AMPure XP. The size-selected ligation product was eluted with 15 μL of 10 mM Tris-HCl (pH 8.0). One microliter of uracil DNA glycosylase (UDG; Enzymatics) was added to the size-selected ligation product. The mixture was incubated at 37 °C for 30 min to exclude the second-strand DNA. For library amplification, 2 μL of the UDG-digested DNA, 1 μL of 2.5 μM index primer[24], 1 μL of 10 μM universal primer[24], 0.5 μL of RNase-free water, and 5 μL of KAPA HiFi HotStart ReadyMix (2×) (KAPA BIOSYSTEMS) were mixed. The DNA fragments with the adapters and an index sequence were amplified using a thermal cycler with the following program: denaturation at 94 °C for 2 min, 16 cycles at 98 °C for 10 s, 65 °C for 30 s, and 72 °C for 30 s for amplification, and 72 °C for 5 min for the final extension. Two rounds of size selection were then performed to remove adapter dimers with equal volume of AMPure XP to the library solution. The purified library was eluted with 10 μL of RNase-free water. The concentration of each library was measured by qPCR using the KAPA SYBR FAST qPCR Master Mix (2×) Kit (KAPA BIOSYSTEMS) and the LightCycler 480 II (Roche Diagnostics, Basel, Switzerland) to mix equal amounts of libraries. One microliter of the mixed library was used for electrophoresis with the Bioanalyzer 2100 and the Agilent High Sensitivity DNA kit (Agilent Technologies, Santa Clara, CA, USA) to check for quality. Sequencing of 50-bp single-ends using HiSeq 2500 (Illumina, San Diego, CA, USA) was carried out by

Macrogen (Seoul, South Korea).

Calculation and normalization of RNA-Seq count data. All obtained reads were processed with Trimmomatic (version 0.33)[26] using the following parameters: TOPHRED33 ILLUMINACLIP:TruSeq3-SE.fa:2:30:10 LEADING:19 TRAILING:19 SLIDINGWINDOW:30:20 AVGQUAL:20 MINLEN:40. This procedure removed adapter sequences (ILLUMINACLIP:TruSeq3-PE.fa:2:30:10), as well as leading and trailing low quality (Q < 19) or N bases (LEADING:19 TRAILING:19). It also trimmed the reads when the average quality per base dropped below 20 with a 30-base wide sliding window (SLIDINGWINDOW:30:20). Thus, trimmed reads with length > 39 nucleotides and average Q score > 19 were obtained. These trimmed reads were then mapped to the reference sequences of IRGSP-1.0_transcript[27,28], rRNAs, and transcripts coded in the mitochondria and chloroplast genomes (NC_001320.1, NC_011033.1) using RSEM (version 1.3.0)[29] and Bowtie (version 1.1.2)[30] with default parameters. Expected counts of each gene in the RSEM outputs were used in R (version 3.4.2)[31] for the following analysis. According to a previous study[32], we removed the effects of misassigned reads in multiplex sequencing experiments with HiSeq 2500 as:

$$c_{i,m,n} = \begin{cases} \tilde{c}_{i,m,n} - 0.0005 \times C_{i,n} \ (\tilde{c}_{i,m,n} > 0.0005 \times C_{i,n}), \\ 0 \ (\tilde{c}_{i,m,n} \leq 0.0005 \times C_{i,n}), \end{cases} \quad (1)$$

where $c_{i.m,n}$ and $\tilde{c}_{i,m,n}$ denote corrected and raw read counts of gene i in sample m in

RNA-Seq library pool n, respectively. $C_{i,n}$ denotes the sum of read counts of gene i in RNA-Seq library pool n.

Then, read counts of gene i in sample m in different libraries were merged, followed by read count normalization based on total read counts of transcripts except for the antisense-oligo targets (reads per million: rpm):

$$r_{i,m} = \frac{c_{i,m}}{C_m} \times 10^6, \tag{2}$$

$$y_{i,m} = \log_2\left(r_{i,m} + 0.1\right), \tag{3}$$

where, $c_{i,m}$, $r_{i,m}$, and $y_{i,m}$ denote merged corrected read counts, normalized read counts, and the log2-transformed value of read counts of gene i in sample m, respectively. C_m denotes total read counts of transcripts except for the antisense-oligo targets in sample m.

Of the 926 samples from 2015 and 143 samples from 2016, we used 887 and 139 RNA-Seq transcriptome datasets, respectively, with more than 10^5 total read counts for all genes, except for the targets of selective depletion (Supplementary Fig. 3a). We then filtered out rarely detected genes (number of samples with read count > 0; ≤ 20% of all samples in the 2015 dataset) from the following analyses to focus on the 23,924 expressed genes only (Supplementary Fig. 3b).

Identification of specific 'Koshihikari' and 'Takanari' single nucleotide polymorphisms. The paired-end reads of 'Koshihikari' (SRR1630927) and 'Takanari'

325 (DRR018360) genomes were trimmed with Trimmomatic (version 0.33)[26] using the
326 parameters stated in the previous section, except for TruSeq3-PE.fa for adapter trimming.
327 The reads were then mapped to the reference genome (IRGSP-1.0, plastid: NC_001320.1,
328 mitochondrion: NC_011033.1)[27] using Bowtie2 (version 2.2.9) with option -N1, which
329 allows a mismatch in seed alignment [33]. By using SAMtools (version 1.3.1)[34], the reads
330 mapped at multiple loci and the duplicated reads, except for each representative read,
331 were discarded. Picard (version 2.1.0) (https://broadinstitute.github.io/picard/) was used
332 to add read group information to each SAM file. All SAM files of each sample were
333 merged and a VCF file was generated using SAMtools (version 1.3.1)[34] and BCFtools
334 (version 1.1)[34]. Single nucleotide polymorphisms (SNPs) were filtered using VCFtools
335 (version 0.1.14)[35] with option --minDP5, which indicates that all alleles supported by
336 more than four reads are reported. The VCF file generated was analyzed using R (version
337 3.4.2)[31] and the R package "vcfR" (version 1.5.0)[36]. We selected 15,117 homozygous
338 SNPs distinguishing 'Koshihikari' and 'Takanari' genotypes for the following analysis of
339 genotype confirmation.
340 **Genotype confirmation**. In previous studies, the genotypes of the CSSLs and BILs were
341 examined based on the genotypes provided by 141 SSR markers[14]. To confirm the
342 genotype of each sample, we estimated each SSR marker genotype based on the SNPs

343 obtained by RNA-Seq. First, to type 'Koshihikari'- or 'Takanari'-type SNPs in each
344 RNA-Seq sample, the trimmed RNA-Seq reads of each sample were mapped to the
345 reference genome as described in the previous section using Bowtie2 (version 2.2.9).
346 Then, the SNPs were filtered using VCFtools (version 0.1.14)[35] with option --minDP3,
347 which indicates that all alleles supported by more than two reads are reported. To estimate
348 each SSR marker genotype in our samples, 'Koshihikari'- and 'Takanari'-type SNPs were
349 counted from the SSR marker position until one of the counts reached "7" (Supplementary
350 Fig. 4a,b). In the cases where the SNPs around an SSR marker position were too sparse,
351 all the SNPs in ±1.5 Mbp from each SSR marker were counted (Supplementary Fig. 4a,c).
352 Then, each estimated SSR marker was categorized into 'Koshihikari'-type, 'Takanari'-
353 type, or undetermined based on most of its SNP counts (Supplementary Fig. 4a-c). Any
354 samples of the estimated genotypes not perfectly matching the SL1201, SL1229, or
355 SL1321 genotypes might be due to the low density of SNPs around the substituted SSR
356 markers or to doubtful SNP calls. For the samples with estimated genotypes similar to
357 one of these lines, imputation of estimated SSR marker genotypes was conducted
358 (Supplementary Fig. 4d). Finally, each sample was labeled with the genotype of the most
359 similar SSR marker composition. Thus, 865 samples were confirmed as 'Koshihikari',
360 'Takanari', or CSSLs and 11 samples as BILs in 2015, and 139 samples as 'Koshihikari',

'Takanari', or CSSLs in 2016 (Supplementary Fig. 3d).

Correlation plot of transcriptomes. Pearson's correlation coefficients of transcriptome data for all pairwise comparisons of the 854 samples from 2015 were calculated as:

$$\rho_{m,n} = \frac{\sum_i (y_{i,m} - \bar{y}_m)(y_{i,n} - \bar{y}_n)}{\sqrt{\sum_i (y_{i,m} - \bar{y}_m)^2}\sqrt{\sum_i (y_{i,n} - \bar{y}_n)^2}}, \quad (4)$$

where $\rho_{m,n}$ denotes the Pearson's correlation coefficient between samples m and n. Mean log2-transformed rpms are denoted as $\bar{y}_m$ and $\bar{y}_n$, respectively.

The heatmap representing the correlations in the time-series order was drawn with the "image.plot" function in the R package "fields" (version 9.0)[37].

Meteorological data. Data of average air temperature per 10 min at the Hirakata Weather Station (34°48′5″N, 135°33′6″E) in 2015 for Takatsuki and at Kyotanabe Weather Station (34°49′8″N, 135°45′6″E) in 2016 for Kizugawa were obtained from the Japan Meteorological Agency. Data of air temperature per minute for the FIT (version 0.0.4)[10] library were prepared by linear interpolation of the data per 10 min with the "approxfun" function in R (version 3.4.2)[31]. Data of global solar radiation per minute at the Osaka Weather Station (34°40′9″N, 135°37′51″E) in 2015 for Takatsuki and at the Nara Weather Station (34°40′4″N, 135°50′2″E) in 2016 for Kizugawa were also obtained from the Japan Meteorological Agency.

Conversion of age to scaled age. To consider the difference in plant's age at heading

379 among genotypes, we defined scaled age. Under the assumption that rice ages linearly

380 before heading at various speeds among genotypes, flowers at scaled age 85, and ages at

381 the same speed after flowering, the scaled age of sample *m* was defined as:

382 $$a_m = \begin{cases} 85\dfrac{d_m - d_m^S}{d_m^F - d_m^S} & (d_m \leq d_m^F), \\ d_m - d_m^F + 85 \; (d_m > d_m^F), \end{cases} \quad (5)$$

383 where d_m is the sampling date, d_m^S is the seeding date, and d_m^F is the flowering date

384 of each line with genotype and seeding date identical to that of sample *m*.

385 **Calculation of precision weights for each observed normalized read count (log2rpm).**

386 Precision weights were utilized in the training step of FIT (version 0.0.4)[10] to deal with

387 over-dispersion of RNA-Seq data[10]. Precision weight for each observed log2rpm value

388 was calculated based on the residuals from the smoothed time-series as in the voom

389 method[10,38] (Supplementary Fig. 14). First, for each gene, spline interpolation against

390 time and observed log2rpm was conducted with the "splinefun" function in R, to infer the

391 function value $s_{iG}(t)$ for interpolation of observed log2rpm of gene *i* at time *t* in $G \in$

392 $\{\mathrm{K}, \mathrm{T}\}$, where K and T denote 'Koshihikari' and 'Takanari'-background genotypes,

393 respectively (Supplementary Fig. 14a). Then, residual standard deviation of gene *i* in *G*-

394 background genotype $\sigma_{i,G}$ was calculated as:

395 $$\sigma_{i,G} = \sqrt{\frac{1}{|M_G|} \sum_{m \in M_G} \left(s_{i,G}(t_m) - y_{i,m}\right)^2}, \quad (6)$$

396 where t_m denotes sampling time of sample *m*. M_G and $|M_G|$ denote indices of samples

397 whose background genotypes are *G* and its size, respectively.

398 The mean log2-transformed read count of gene *i* in *G*-background genotype was then

399 calculated as:

400 $$\bar{c}_{i,G} = \frac{1}{|M_G|}\sum_{\mathrm{m}\in\mathrm{M_G}} y_{i,m} + \frac{1}{|M|}\sum_{m\in M} \log_2 C_m + 1 - \log_2 10^6\,, \quad (7)$$

401 where *M* denotes sample indices in both background genotypes.

402 A LOWESS curve[39] was then fitted to the square-root of residual standard deviations as

403 a function of mean log2-transformed read counts, resulting in *l*(), which represents the

404 trend between log2-transformed read counts and square-root of residual standard

405 deviations (Supplementary Fig. 14b).

406 Similar to $\bar{c}_{i,G}$, smoothed log2rpm of gene *i* in sample *m* was converted to

407 smoothed log2-transformed read counts of gene *i* in sample *m* as:

408 $$\hat{c}_{i,m} = s_{i,G_m}(t_m) + \log_2 C_m + 1 - \log_2 10^6\,, \quad (8)$$

409 where G_m denotes the background genotype of sample *m*.

410 Finally, precision weight for observed log2rpm of gene *i* in sample *m* was calculated as

411 (Supplementary Fig. 14c):

412 $$w_{i,m} = \frac{1}{l\left(\hat{c}_{i,m}\right)^4}. \quad (9)$$

413 **Prediction model development for 'Koshihikari' and 'Takanari'**. A prediction model

414 for gene expression dynamics in 'Koshihikari' and 'Takanari' was developed with the R
415 package "FIT" (version 0.0.4)[10]. In this section, we describe how "FIT" developed a
416 prediction model.

417 There are several strategies for the integrative analysis of the transcriptome and
418 meteorological data to investigate environmental responses in field conditions[40–44]. In the
419 FIT package, we employed the statistical modeling approach for a comprehensive
420 relationship among these data and for understanding the effects of age and circadian clock
421 of plants. We used the transcriptome data in combination with the meteorological data,
422 namely air temperature and global solar radiation, which were measured at a
423 meteorological station near the field from where the samples were acquired. Because
424 there was no remarkable disease or pest damage during cultivation and no difference in
425 fertilization among samples, the model did not include these effects.

426 The log2rpm of the normalized expression for gene *i* of sample *m* was described by a
427 simple linear model:

428 $$y_{i,m} = \alpha_i + \boldsymbol{x}_{i,m}^T \boldsymbol{\beta}_i + \epsilon_{i,m}, \quad (10)$$

429 where α_i, $\boldsymbol{x}_{i,m}$, and $\boldsymbol{\beta}_i$ are a constant term, a vector of eight explanatory variables, and
430 regression coefficients, respectively. The third term $\varepsilon_{i,m}$ is the independently and
431 identically distributed noise drawn from a Gaussian distribution. The explanatory

variables were plant's scaled age, circadian clock, optional variable, response to environmental stimuli, and the interactions between scaled age and circadian clock and between scaled age and response to environmental stimuli. The optional variable was not used in this study. The plant's scaled age corresponded to the scaled number of days after transplanting. See section "Conversion of age to scaled age" for the definition of scaled age. The values of the scaled age were normalized to mean 0 and variance 1. The circadian clock, with the peak at an arbitrary time, was represented by two variables corresponding to its cosine and sine components:

$$\begin{aligned} &\cos\left(\frac{2\pi t_m}{24}\right), \\ &\sin\left(\frac{2\pi t_m}{24}\right), \end{aligned} \tag{11}$$

where t_m is the time at which sample m was obtained. Similarly, the interaction between scaled age and circadian clock of a plant was represented by two variables.

Although we employed a linear model for simplicity, the actual gene expression can exhibit more complex responses. The complexity of fluctuations in gene expression is mainly explained by the nonlinearity of the response to the environmental stimuli, which is the cumulative sum of the nonlinearly transformed environmental stimuli for a given period. Such cumulative sum of temperature, for example, is designated as cumulative temperature and it is widely used for the prediction of reproduction timing and vegetation

449 in ecology and of harvest timing in agriculture[45,46]. By applying this to gene expression,

450 the environmental response of gene i in sample m can be described as:

451 $$\hat{e}_{i,m} = \sum_{\tau=t_m-p_i}^{t_m} g_i(\tau) \max\left(0, \tanh\left(\pm \exp\left(\gamma_i^{(f)}\right)\left(e_\tau - \theta_i^{(f)}\right)\right)\right) \times \sqrt{\exp\left(-2\gamma_i^{(f)}\right) + 1}, \quad (12)$$

452 where p_i is the period during which gene expression was affected by an environmental

453 stimulus, e_τ is the value of a meteorological parameter at time τ normalized to mean 0 and

454 variance 1, $\theta_i^{(f)}$ is the response threshold to the stimulus, and $\gamma_i^{(f)}$ is the parameter that

455 controls the shape of the response (Supplementary Fig. 15). Here, the meteorological data

456 consisted of air temperature and global solar radiation. Of these parameters, the one

457 yielding the highest likelihood was selected for each gene in the optimization procedure.

458 In the limit $\gamma_i^{(f)} \to \infty$, a gene responds to an environmental stimulus in a dose-

459 independent manner, that is, the response is constant if a meteorological parameter

460 exceeds the threshold (Supplementary Fig. 15). Conversely, as this parameter approaches

461 minus infinity, the response also approaches the dose-dependent manner (Supplementary

462 Fig. 15). The sign of the value in the hyperbolic function determines whether the gene

463 responds to an environmental stimulus larger than (positive) or smaller than (negative)

464 the threshold. This sign was decided in the optimization procedure based on the parameter

465 likelihood. The last term $\sqrt{\exp\left(-2\gamma_i^{(f)}\right) + 1}$ ensures that the scale of the response is

466 largely independent of this parameter. The function $g_i(\tau)$ is a gate function that explains

the time-of-day-specific environmental responses. A gate function is defined as:

$$g_i(\tau) = n_i\left(\tanh\left(\exp\left(\gamma_i^{(g)}\right)\left(\cos\left(2\pi(\tau - \psi_i)\right)/24 - \theta_i^{(g)}\right)\right)\right), \quad (13)$$

where $\gamma_i^{(g)}$ controls the shape of the gate as $\gamma_i^{(f)}$ in the environmental response, $\theta_i^{(g)}$ determines the opening length of the gate, and $n_i()$ is the function which normalizes an input value $[0,1]$. This gate function approaches a sinusoidal and rectangular wave with the small and large values of $\gamma_i^{(g)}$, respectively. If the value of $\theta_i^{(g)}$ is smaller than -1, the gate is a constant value in the limit $\gamma_i^{(g)} \to \infty$, that is, the gate is always open (Supplementary Fig. 15).

It can be considered that some explanatory variables might contribute to neither the explanation nor the prediction of the expression of many genes. Thus, we simultaneously preformed the optimization and variable selection of the regression coefficients by using group lasso[47]. Let I be the index set for scaled age, environmental response, and the interaction between scaled age and environmental response. The cost function to be minimized is thus defined as:

$$L(\boldsymbol{\beta}) = \sum_{m \in M}\left(\hat{y}_{i,m} - y_{i,m}\right)^2 + \lambda\left(\sum_{k \in I}\zeta_k|\beta_k| + \zeta_c\sqrt{\beta_{cos}^2 + \beta_{sin}^2} + \zeta_{dc}\sqrt{\beta_{dcos}^2 + \beta_{dsin}^2}\right), \quad (14)$$

where λ and $\boldsymbol{\zeta}$ are the regularization parameter and the adaptive weights, respectively.

484 The values of the adaptive weights are decided as in the adaptive lasso[48] or the adaptive
485 group lasso[49]. Further, the adaptive weights were used to absorb differences in the degrees
486 of freedom by multiplying the covariates of the environmental response and interaction
487 between the environmental response and scaled age by seven, as seven free parameters
488 were considered. We selected the value of λ as the largest value for which the cross
489 validation (CV) error was smaller than the sum of the minimum CV error and its standard
490 error.

491 The values of the parameters related to environmental responses, which are
492 $p_i, \theta_i^{(f)}, \gamma_i^{(f)}$, and the sign of the value in the hyperbolic function in Eq. (12), and
493 $\theta_i^{(g)}, \gamma_i^{(g)}$, and ψ_i in Eq. (13), are selected by the Nelder-Mead algorithm[50]. Because the
494 likelihood function is complex and has multiple local maxima, a grid search was
495 performed before optimization. In each step of the optimization by the Nelder-Mead
496 algorithm, we performed an ordinary least square regression instead of group lasso (Eq.
497 (14)) to reduce the computational cost.

498 'Koshihikari' and 'Takanari'-prediction models were developed based on time-
499 series RNA-Seq data of the 'Koshihikari'-background CSSLs and 'Koshihikari' and of
500 the 'Takanari'-background CSSLs and 'Takanari' in 2015, respectively (Supplementary
501 Fig. 2b). Values of the expressed genes (a matrix of log2rpm for each sample), sample

502 attributes (a matrix of sampling year, month, day, hour, and min and scaled age for each
503 sample), weather data (a matrix of air temperature and global solar radiation at every
504 minute in 2015), and precision weights for the matrix of log2rpm of each sample
505 (Supplementary Fig. 16) were prepared. Details on the parameters for 'FIT' are listed in
506 https://github.com/naganolab/Rice_edQTL-analysis_and_edQTL-based-
507 prediction/analysis.R. As a result, the functions $f_{i,G}()$ were yielded for gene i. A log2rpm
508 of gene i of sample m was predicted as (Supplementary Fig. 16):

509 $$\hat{y}_m = f_{i,G_m}(t_m, a_m, \boldsymbol{e}), \tag{15}$$

510 where $\boldsymbol{e}$ denotes meteorological data (air temperature and global solar radiation) at every
511 minute.

512 **Detection of genes with expression dynamics polymorphism.** To increase detection
513 power in multiple tests, we focused on genes with expression dynamics polymorphism.
514 Euclidean distances of gene i between 'Koshihikari'- and 'Takanari'-prediction models
515 (d_i) were calculated as:

516 $$d_i = \sqrt{\sum_{m \in M} \left(f_{i,\mathrm{K}}(t_m, a_m, \boldsymbol{e}) - f_{i,\mathrm{T}}(t_m, a_m, \boldsymbol{e}) \right)^2}. \tag{16}$$

517 As a result, 3696 genes with $d_i > 40$ were defined as genes with expression dynamics
518 polymorphism (Supplementary Fig. 3e).

519 **Detection of edQTL.** According to background genotypes, the sum of residual errors

520 between the observed log2rpm and predicted log2rpm of gene i in the 'Koshihikari'- or

521 'Takanari'-models (E_i) were calculated as (Supplementary Fig. 3a):

522 $$E_{i,m} = \left|\hat{y}_{i,m} - y_{i,m}\right|, \tag{17}$$

523 $$E_i = \sum_{m \in M} E_{i,m}\,, \tag{18}$$

524 where E_{im} denotes the residual error between the observed log2rpm and predicted

525 log2rpm of gene i in sample m on the assumption that the 'Koshihikari'- or 'Takanari'-

526 type gene expression dynamics shown by sample m is determined based on background

527 genotype G_m. Similarly, on the assumption that edQTL for gene i exist around SSR marker

528 g, the sum of residual errors between the observed log2rpm and predicted log2rpm of

529 gene i in the 'Koshihikari'- or 'Takanari'-models $\tilde{E}_{i,g}$ were calculated as

530 (Supplementary Fig. 3a):

531 $$\tilde{E}_{i,m,g} = \left|f_{i,Q_{g,m}}(t_m, a_m, \boldsymbol{e}) - y_{i,m}\right|, \tag{19}$$

532 $$\tilde{E}_{i,g} = \sum_{m \in M} \tilde{E}_{i,m,g}\,, \tag{20}$$

533 where $\tilde{E}_{i,m,g}$ denotes the residual error between the observed log2rpm and predicted

534 log2rpm of gene i in sample m and $Q_{g,m}$ denotes genotype of SSR marker g in sample

535 m. Thus, $f_{i,Q_{gm}}(t_m, a_m, \boldsymbol{e})$ denotes predicted log2rpm in sample m assuming the

536 existence of edQTL for gene i around SSR marker g. The improvement in residual errors

537 assuming that edQTL for gene i exists around g was therefore calculated as

538 (Supplementary Fig. 3a,b):

539 $$I_{i,g} = E_{i,m} - \tilde{E}_{i,m,g}. \quad (21)$$

540 To evaluate the statistical significance of the improvement in the sum of residual
541 errors, 1000 permutations of samples with 'Koshihikari'- and 'Takanari' genetic
542 background were performed (Supplementary Fig. 6e). For each permutation, the
543 improvement in residual errors on the assumption that edQTL for gene *i* exist around SSR
544 marker *g* was calculated (Supplementary Fig. 6b). Then, we generated the null
545 distribution of improvements in residual errors by fitting log-normal distribution to
546 obtained improvements using the fitdist function in the R package "fitdistrplus"[51], with
547 the fitting method of quantile matching of 70% and 1% (Supplementary Fig. 6c). For
548 almost all genes with expression dynamics polymorphism, the 99.9 percentiles of the
549 fitted log-normal distributions were higher than (2863/3696 genes) or almost the same as
550 those of the distribution obtained by the permutations (Supplementary Fig. 6d), indicating
551 that the *p*-value obtained from log-normal distribution was conservative. Thus, the *p*-
552 value for the improvement in the sum of residual errors for gene *i* on the assumption that
553 an edQTL affecting gene expression dynamics of gene *i* exists around SSR marker *g* was
554 obtained based on the null distribution.

555 The adjustment for multiple comparisons against *p*-values was performed with

556 the Benjamini-Hochberg method[52] using the p.adjust function in R. Finally, we defined

557 the markers with the smallest adjusted p-value (< 0.05) in each peak for each gene as the

558 edQTL affecting each gene expression dynamics.

559 **edQTL-based prediction of gene expression dynamics (edQTL model)**. To predict

560 gene expression with edQTL information, the 'Koshihikari'- or 'Takanari'-model was

561 selected based on the genotype of the edQTL for edQTL-affected genes and of genetic

562 backgrounds for the other expressed genes. Let $\mathcal{G}_{\mathrm{i}}$ denote the set of edQTLs affecting

563 gene i. We selected the model for prediction of log2rpm of gene i in sample m as:

564 $$g_{i,m} = \begin{cases} \tilde{G}_m, & (\exists g \in \mathcal{G}_i, Q_{g,m} \neq G_m), \\ G_m, & \text{otherwise}, \end{cases} \tag{22}$$

565 where $\tilde{\mathrm{G}}_m$ denotes the opposite genotype to that of the genetic background in sample m,

566 that is, $\tilde{G}_m = \mathrm{K}$, if sample m is 'Takanari'-background, and vice versa. Then, predicted

567 log2rpm of gene i in sample m was calculated as:

568 $$\hat{y}_{i,m}^{\mathrm{ED}} = f_{i,g_{i,m}}(t_m, a_m, \boldsymbol{e}). \tag{23}$$

569 **Statistical test for the accuracy of gene expression dynamics prediction in BILs**. The

570 sums of prediction errors of all genes affected by each edQTL for the BILs based on the

571 'Koshihikari' model (E_{K}), 'Takanari' model (E_{T}), and edQTL model (E^{ED}) were

572 calculated as:

573 $$E_G = \sum_{i \in I^{\mathrm{ED}}, m \in M^{\mathrm{B}}} \left| f_{i,G}(t_m, a_m, \boldsymbol{e}) - y_{i,m} \right|, \tag{24}$$

$$E^{\mathrm{ED}} = \sum_{i \in I^{\mathrm{ED}}, m \in M^{\mathrm{B}}} \left| \hat{y}_{i,m}^{\mathrm{ED}} - y_{i,m} \right|, \tag{25}$$

where M^{B} denotes the indices of BIL samples and I^{ED} denotes the genes affected by the edQTL.

To evaluate the statistical significance of improvement in the sums of prediction errors, 10,000 permutations were performed for the markers in each BIL genome. In each permutation, the sums of prediction errors of all genes affected by the edQTL for the BILs were calculated with the edQTL model. Then, the value corresponding to percentile 0.1 of the sums of prediction errors was calculated.

Code availability. The scripts used in this study are available from https://github.com/naganolab/Rice_edQTL-analysis_and_edQTL-based-prediction.

Data availability. All datasets generated and/or used in this study are available from PRJDB7234.

1. Poorter, H. *et al.* Pampered inside, pestered outside? Differences and similarities between plants growing in controlled conditions and in the field. *New Phytol.* **212**, 838–855 (2016).

2. Brachi, B., Morris, G. P. & Borevitz, J. O. Genome-wide association studies in plants: the missing heritability is in the field. *Genome Biol.* **12**, 232 (2011).

3. Eichler, E. E. *et al.* Missing heritability and strategies for finding the underlying causes of complex disease. *Nat. Rev. Genet.* **11**, 446–450 (2010).

4. Wang, E. *et al.* The uncertainty of crop yield projections is reduced by improved temperature response functions. *Nat. Plants* **3**, 17102 (2017).

5. Horiuchi, Y. *et al.* Global expression differences and tissue specific expression differences in rice evolution result in two contrasting types of differentially expressed genes. *BMC Genomics* **1099**, 16 (2015).

6. Wang, J. *et al.* An expression quantitative trait loci-guided co-expression analysis for constructing regulatory network using a rice recombinant inbred line population. *J. Exp. Bot.* **65**, 1069–1079 (2014).

7. Jansen, R. C. & Nap, J. P. Genetical genomics: The added value from

segregation. *Trends Genet.* **17**, 388–391 (2001).

8. Wang, J. *et al.* A global analysis of QTLs for expression variations in rice shoots at the early seedling stage. *Plant J.* **63**, 1063–1074 (2010).

9. Kuroha, T. *et al.* eQTLs Regulating Transcript Variations Associated with Rapid Internode Elongation in Deepwater Rice. *Front. Plant Sci.* **8**, 1753 (2017).

10. Iwayama, K., Aisaka, Y., Kutsuna, N. & Nagano, A. J. FIT: Statistical modeling tool for transcriptome dynamics under fluctuating field conditions. *Bioinformatics* **33**, 1672–1680 (2017).

11. Nagano, A. J. *et al.* Deciphering and prediction of transcriptome dynamics under fluctuating field conditions. *Cell* **151**, 1358–1369 (2012).

12. Matsuzaki, J., Kawahara, Y. & Izawa, T. Punctual transcriptional regulation by the rice circadian clock under fluctuating field conditions. *Plant Cell* **27**, 633–648 (2015).

13. Ookawa, T. *et al.* Precise estimation of genomic regions controlling lodging resistance using a set of reciprocal chromosome segment substitution lines in rice. *Sci. Rep.* **30572**, 6 (2016).

14. Takai, T. *et al.* Genetic mechanisms underlying yield potential in the rice high-yielding cultivar Takanari, based on reciprocal chromosome segment substitution lines. *BMC Plant Biol.* **295**, 14 (2014).

15. Adachi, S. *et al.* The mesophyll anatomy enhancing CO2diffusion is a key trait for improving rice photosynthesis. *J. Exp. Bot.* **64**, 1061–1072 (2013).

16. Adachi, S. *et al.* Genetic architecture of leaf photosynthesis in rice revealed by different types of reciprocal mapping populations. *J. Exp. Bot.* (2019).

17. Sakamoto, T. An Overview of Gibberellin Metabolism Enzyme Genes and Their Related Mutants in Rice. *Plant Physiol.* **134**, 1642–1653 (2004).

18. Sultan, S. E. Phenotypic plasticity and plant adaptation. *Acta Bot. Neerl* **44**, 363–383 (1995).

19. Fournier-Level, A. *et al.* Predicting the evolutionary dynamics of seasonal adaptation to novel climates in Arabidopsis thaliana. *Proc. Natl. Acad. Sci. U. S. A.* **113**, E2812-21 (2016).

20. Skelly, D. A., Ronald, J. & Akey, J. M. Inherited Variation in Gene Expression. *Annu. Rev. Genomics Hum. Genet.* **10**, 313–332 (2009).

21. Kaur, J., Rahat, B., Thakur, S. & Kaur, J. Trends in Precision Medicine. *Prog. Challenges Precis. Med.* 269–299 (2017). doi:10.1016/B978-0-12-809411-2.00015-5

22. Kremling, K. A. G. G. *et al.* Dysregulation of expression correlates with rare-allele burden and fitness loss in maize. *Nature* **555**, 520–523 (2018).

23. Horinouchi, T. *et al.* Prediction of cross-resistance and collateral sensitivity by gene expression profiles and genomic mutations. *Sci. Rep.* **14009**, 7 (2017).

24. Nagano, A. J., Honjo, M. N., Mihara, M., Sato, M. & Kudoh, H. Detection of plant viruses in natural environments by using RNA-Seq. *Plant Virol. Protoc. Methods Mol. Biol.* **1236**, 89–98 (2015).

25. Morlan, J. D., Qu, K. & Sinicropi, D. V. Selective depletion of rRNA enables whole transcriptome profiling of archival fixed tissue. *PLoS One* **7**, e42882 (2012).

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

27. Kawahara, Y. *et al.* Improvement of the Oryza sativa Nipponbare reference

genome using next generation sequence and optical map data. *Rice* **6**, 4 (2013).

28. Sakai, H. *et al.* Rice annotation project database (RAP-DB): an integrative and interactive database for rice genomics. *Plant Cell Physiol.* **54**, 1–11 (2018).

29. Li, B. & Dewey, C. N. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* **12**, 323 (2011).

30. Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **10**, R25 (2009).

31. R Core Team. R: A language and environment for statistical computing. R Foundation for Statis-tical Computing, Vienna, Austria. (2017).

32. Kamitani, M., Nagano, A. J., Honjo, M. N. & Kudoh, H. RNA-Seq reveals virus-virus and virus-plant interactions in nature. *FEMS Microbiol. Ecol.* **92**, 1–11 (2016).

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

34. Li, H. *et al.* The sequence alignment / map (sam) Format and SAMtools.

Bioinformatics **25**, 2078–2079 (2009).

35. Danecek, P. *et al.* The variant call format and VCFtools. *Bioinformatics* **27**, 2156–2158 (2011).

36. Knaus, B. J. & Grünwald, N. J. vcfr: a package to manipulate and visualize variant call format data in R. *Mol. Ecol. Resour.* **17**, 44–53 (2017).

37. Nychka, D., Furrer, R., Paige, J. & Sain, S. fields: Tools for spatial data. (2017). doi:10.5065/D6W957CT

38. Law, C. W., Chen, Y., Shi, W. & Smyth, G. K. voom: precision weights unlock linear model analysis tools for RNA-seq read counts. *Genome Biol.* **15**, R29 (2014).

39. Cleveland, W. S. LOWESS: A Program for Smoothing Scatterplots by Robust Locally Weighted Regression. *Am. Stat.* **35**, 54 (1981).

40. Wilkins, O. *et al.* EGRINs (Environmental gene regulatory influence networks) in rice that function in the response to water deficit, high temperature, and agricultural environments. *Plant Cell* **28**, 2365–2384 (2016).

41. Cronn, R. *et al.* Transcription through the eye of a needle: Daily and annual

cyclic gene expression variation in Douglas-fir needles. *BMC Genomics* **18**, (2017).

42. Richards, C. L., Rosas, U., Banta, J., Bhambhra, N. & Purugganan, M. D. Genome-wide patterns of Arabidopsis gene expression in nature. *PLoS Genet.* **8**, (2012).

43. Izawa, T. *et al.* Os- *GIGANTEA* Confers Robust Diurnal Rhythms on the Global Transcriptome of Rice in the Field. *Plant Cell* **23**, 1741–1755 (2011).

44. Sjödin, A., Wissel, K., Bylesjö, M., Trygg, J. & Jansson, S. Global expression profiling in leaves of free-growing aspen. *BMC Plant Biol.* **8**, (2008).

45. Kira, T. On the altitudinal arrangement of climatic zones in Japan—a contribution to the rational land utilization in cool highlands. *Kanti-Nogaku* **2**, 143–173 (1948).

46. Schwartz, M. *Phenology : an integrative environmental science.* (Springer, 2014).

47. Yuan, M., Yuan, M. & Lin, Y. Model selection and estimation in regression with grouped variables. *J. R. Stat. Soc. Ser. B* **68**, 49--67 (2006).

48. Zou, H. The adaptive lasso and its oracle properties. *J. Am. Stat. Assoc.* **101**, 1418–1429 (2006).

49. Wang, H. & Leng, C. A note on adaptive group lasso. *Comput. Stat. Data Anal.* **52**, 5277–5286 (2008).

50. Nelder, J. A. & Mead, R. A Simplex Method for Function Minimization. *Comput. J.* **7**, 308–313 (1965).

51. Delignette-Muller, M. L. & Dutang, C. fitdistrplus : An R Package for Fitting Distributions. *J. Stat. Softw.* **64**, 1–34 (2015).

52. Benjamini, Y. & Hochberg, Y. Controlling the False Discovery Rate : A Practical and Powerful Approach to Multiple Testing. *J. R. Stat. Soc. Ser. B* **57**, 289–300 (1995).

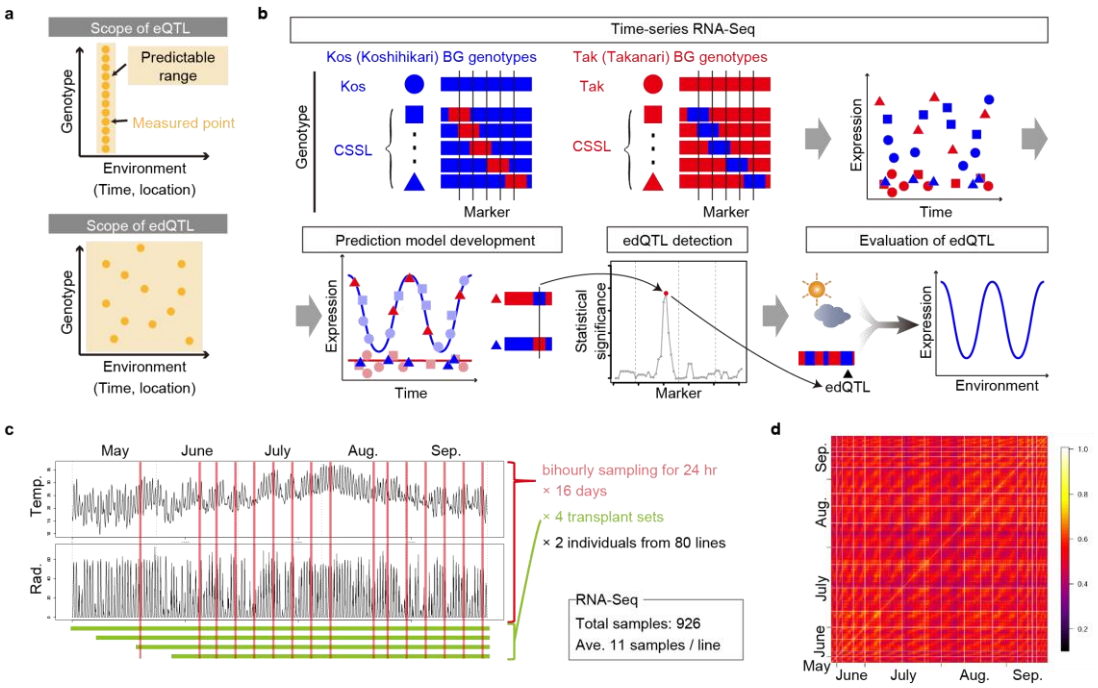
Supplementary Information is available in the online version of the paper.

Acknowledgments. We thank the field work support team for help at field sites, Fumie Kobayashi for assistance with laboratory experiments, and Yasuhiro Sato for valuable discussions. We thank Toshio Yamamoto (National Agriculture and Food Research Organization) for providing seeds of the CSSLs and BILs between 'Koshihikari' and 'Takanari'. This work was supported by the JST CREST JPMJCR15O2 and JSPS KAKENHI Grants JP16H06171 and JP16H01473 to A.J.N.

Author contributions. M.K. analyzed the data. R.L.S., H.S., S.O. and S. A. performed field work. M.K., A.T., A.D., Y.H., and Y.K. conducted RNA-Seq experiments. K.I. developed the FIT library. A.J.N. designed the study. M.K. and A.J.N. wrote the manuscript with input from all co-authors.

Declaration of interests. The authors declare no competing interests.

728 **Figures**



729

730 **Figure 1 | Concept and workflow of expression dynamics quantitative trait loci**

731 **(edQTL) detection. a**, Conceptual differences in the scopes of expression quantitative

732 trait loci (eQTL) and edQTL. **b**, Workflow of edQTL detection and its evaluation using

733 chromosome segment substitution lines (CSSLs). **c**, Summary of the sampling design for

734 edQTL detection. The left panels show plots of meteorological data [air temperature

735 (Temp., °C) and global solar radiation (Rad., kJ m^{-2} min^{-1})] in Takatsuki from May to

736 September in 2015. Vertical red lines represent the sampling time points. **d**, Pearson's

737 correlations among the 854 samples used for developing the prediction model based on

738 expression data of 23,294 expressed genes. White lines indicate the border of each

bihourly sampling set.

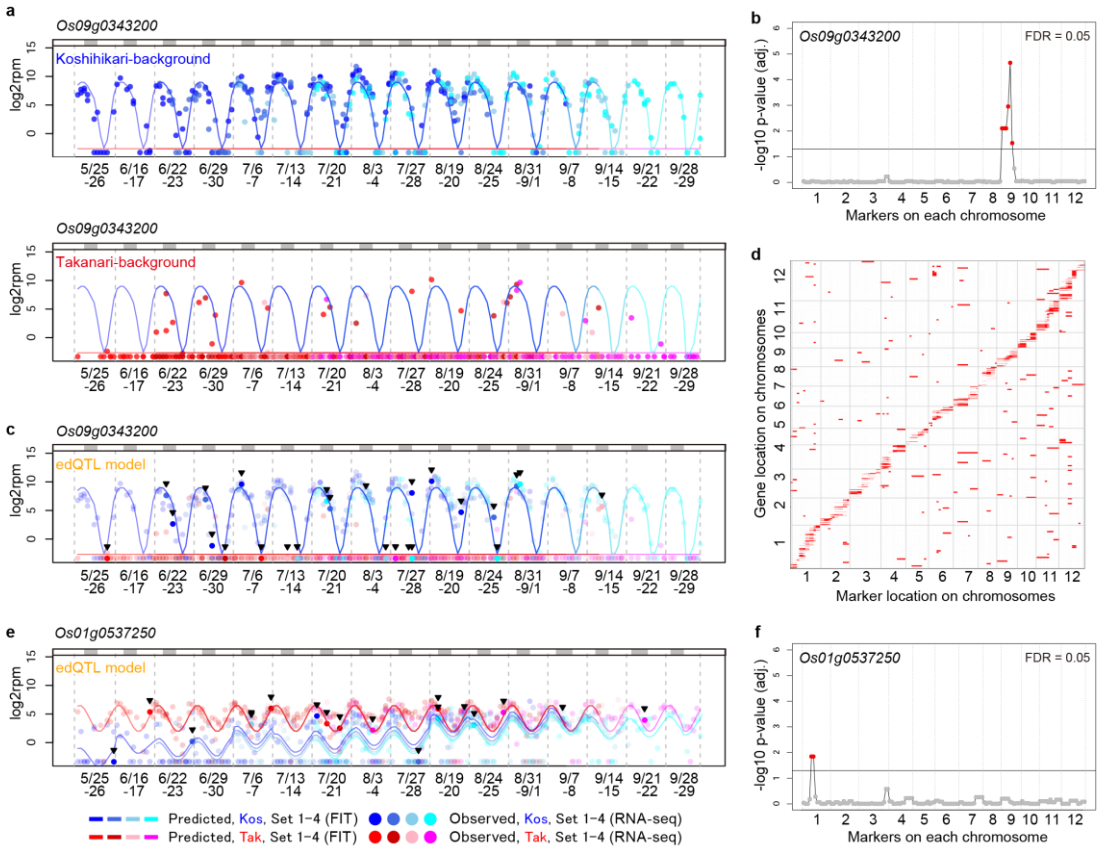


Figure 2 | Expression dynamics quantitative trait loci (edQTL) detection revealed *cis*- and *trans*-edQTL that were involved in environmental responses. a, c, e, Observed and predicted expressions (log_2rpm) of *Os09g0343200* (**a,c**) and *Os01g0537250* (**e**) in Takatsuki in 2015. The predicted expressions were calculated using 'FIT' based on scaled age and environmental information (time, air temperature, and global solar radiation).Blue and red/pink lines indicate predicted expression levels in 'Koshihikari' and 'Takanari' in transplant sets 1, 2, 3, and 4, respectively. Blue and red/pink points indicate the expression level obtained by RNA-sequencing for samples in transplant sets 1, 2, 3, and 4 of individuals with 'Koshihikari'-background genotype (BG)

750 and ‘Takanari’-BG in (**a**) or ‘Koshihikari’- and ‘Takanari’-type edQTL in (**c,e**),

751 respectively. In (**c,e**), strong colored points emphasized by the arrowheads indicate

752 samples harboring edQTL different from their BGs, and light-colored points indicate the

753 samples harboring edQTL identical to their BGs. The upper gray bar indicates dark

754 periods (global solar radiation < 0.3 kJ m^{-2} min^{-1}). **b,f**, edQTL regulating *Os09g0343200*

755 (**b**) and *Os01g0537250* (**f**). **d**, Position of the 1675 edQTL are shown as red bars (false

756 discovery rate = 0.05). X-axis and Y-axis represent the positions of markers with edQTL

757 and the positions of genes influenced by edQTL, respectively.

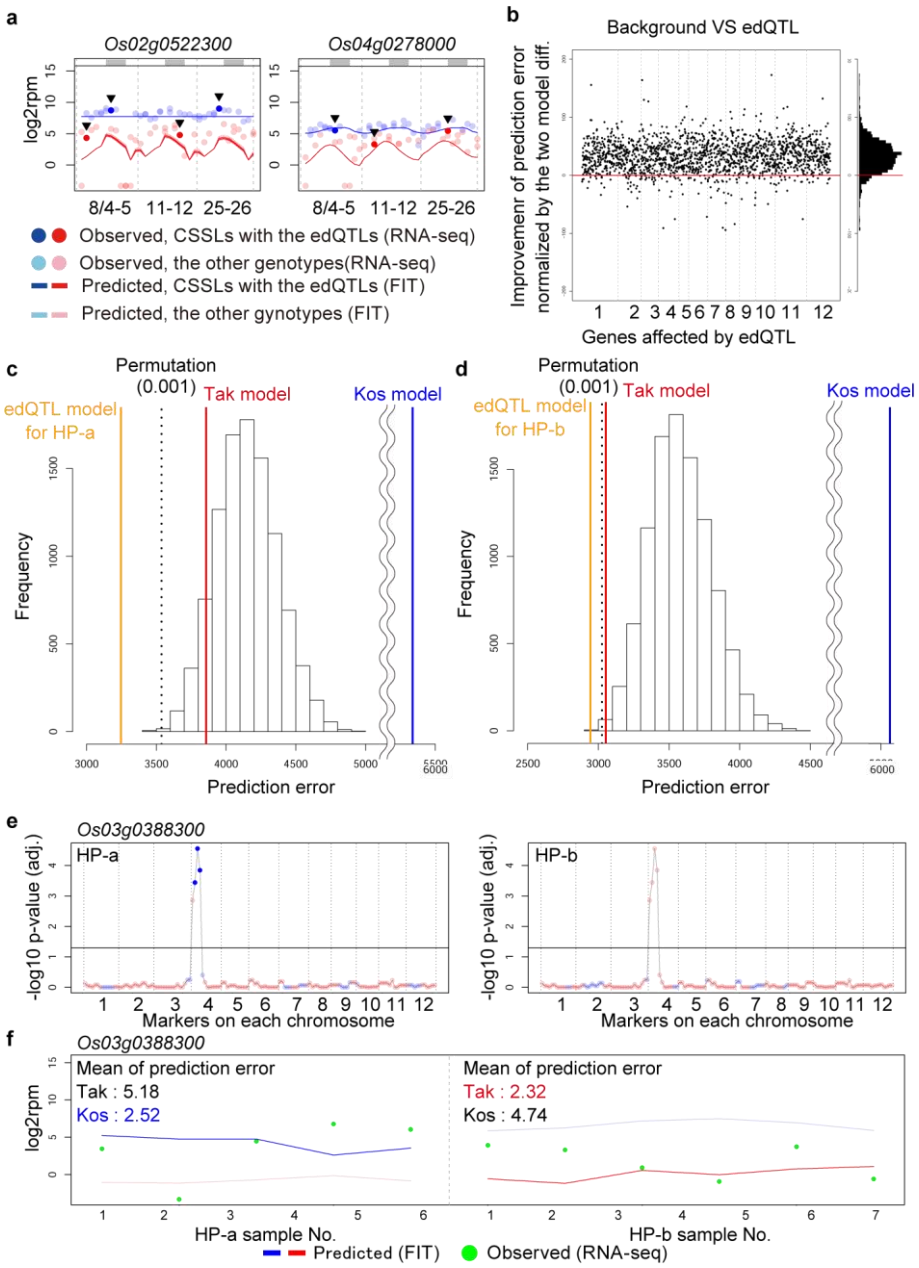


Figure 3 | Expression dynamics quantitative trait loci (edQTL)-based prediction of gene expression dynamics under different environments and in cultivars with more complex genotypes than the chromosome segment substitution lines (CSSLs). a, Examples of prediction of expression dynamics in Kizugawa in 2016 based on

763 environmental information and edQTL in transplant set 1. Points in intense colors
764 emphasized by arrowheads indicate samples harboring edQTL different from their
765 background genotypes and light colors indicate samples harboring edQTL identical to
766 their background genotypes. The upper gray bars indicate dark periods (global solar
767 radiation < 0.3 kJ m^{-2} min^{-1}). **b,** Effects of edQTL over the prediction of gene expression
768 dynamics in Kizugawa in 2016. Genes influenced by edQTL are in order of position on
769 the chromosomes along the horizontal axis. **c,d,** Prediction accuracy of edQTL model for
770 HP-a (**c**) and HP-b (**d**). The blue, red, and orange vertical lines indicate the sums of
771 prediction errors based on 'Koshihikari', 'Takanari', and edQTL models. The histogram
772 shows the distribution of the sums of prediction errors based on the edQTL model in
773 10,000 permutations of markers in HP-a or HP-b genomes. The dashed vertical line
774 indicates the 0.1% percentile of the distribution. **e,** edQTL for *Os03g0388300* and
775 genotypes of HP-a and HP-b. Dark blue points indicate significant 'Koshihikari'-type
776 markers. **f,** Prediction of *Os03g0388300* expression in HP-a and HP-b. Intense color lines
777 are applied models for HP-a and HP-b.
778

779 **Supplemental Tables**

780 **Table S1 SSR marker composition in the rice lines used in the present study**

781 "Position" indicates the position of each SSR marker on the chromosomes in the
782 IRGSP1.0 reference genome. A and B indicate 'Koshihikari'- and 'Takanari'-type alleles,
783 respectively.

784

785 **Table S2 Sample attributes used in this study**

786 "TransPlantSet" indicates the transplant sets in each year. Details of seeding and
787 transplant dates in each set are described in the methods. "Year", "month", "day", and
788 "hour" columns indicate the sampling date. The "LineName" column indicates the
789 predicted genotypes based on SNPs detected in each RNA-Seq. The "field" column
790 indicates the location of the fields.

791

792 **Table S3 *Oryza sativa* genes and their expressions**

793 The "Expression" column indicates whether the genes are "expressed genes" or not. "Yes"
794 indicates that "expressed genes" were detected in more than 20% of all samples. The
795 "Polymorphism" column indicates whether the genes show expression dynamics
796 polymorphism between 'Koshihikari' and 'Takanari' or not. "Yes" indicates that the

797 Euclidian distances of predicted expression dynamics between ‘Koshihikari’ and

798 ‘Takanari’ were more than 40. The "edQTL", "cis.edQTL", and "trans.edQTL" columns

799 indicate whether the genes have the designated edQTL or not.